



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

지리학석사학위논문

SpVAR-Lasso 모델링을 통한
주택 매매-전세가격의 시공간적
상호작용 패턴 분석

2016년 2월

서울대학교 대학원

지 리 학 과

진 찬 우

SpVAR-Lasso 모델링을 통한 주택 매매-전세가격의 시공간적 상호작용 패턴 분석

지도교수 이 건 학

이 논문을 지리학석사 학위논문으로 제출함
2015년 10월

서울대학교 대학원
지 리 학 과
진 찬 우

진찬우의 지리학석사학위논문을 인준함
2016년 1월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

| 국문초록 |

본 연구에서는 주택 매매가와 전세가 간의 시공간적 상호작용 패턴을 파악하기 위한 새로운 모델을 제시하고, 이를 실제 수도권 주택 시장 분석에 활용함으로써 새로운 방법론의 적용 가능성을 평가하였다. 기존의 시공간 모델인 SpVAR은 과소표본에 의한 자유도 문제로 세분화된 주택 가격 간의 국지적인 상호작용 패턴을 포착하기 어려웠기 때문이다. 이에 따라 본 연구에서는 과소표본 문제로부터 자유로운 Lasso를 SpVAR 모델의 추정 방법으로 제시함(SpVAR-Lasso)으로써 이러한 문제를 해결하고자 하였으며, 이를 주택 가격 간 발생하는 시공간적 상호작용을 분석하는 도구로 활용하는 방안을 제시하였다. 자료의 시공간적 특성 탐색 - 모델링 - 적용 및 평가로 이어지는 SpVAR-Lasso 모델의 적용 가능성 평가는 수도권 65개 시군구의 최근 5년간 주택 종류별 매매가와 전세가를 대상으로 실시하였다. 이에 따른 분석 결과를 요약하면 다음과 같다.

탐색적 접근에서는 SpVAR을 적용하기 위한 주택 가격 자료의 시공간적 효과를 측정하고 평가하였다. 시계열적 정상성을 검토한 결과 비정상성이 입증되어 이를 제거하고자 1차 차분한 가격변화량을 분석 대상으로 삼았다. 이러한 주택 가격의 변화량에서 공간적 의존성과 공간적 이질성이 모두 확인되었다. 특히 아파트 전세가격은 상당히 높게 공간적으로 상호 의존적인 것으로 나타났고, 주변 지역의 다른 변수와도 밀접하게 연관이 있음이 탐지되었다. 또한 변수 간 관계가 지역에 따라 다르게 나타나는 공간적 이질성 역시 포착되었다. 그러므로 주택 가격 간 상호작용 패턴을 분석하기 위해서는 시공간적 의존성을 필터링하면서 공간적 이질성을 효과적으로 다룰 수 있는 SpVAR-Lasso가 적합한 방법인 것으로 나타났다.

이에 따라 SpVAR-Lasso를 활용하여 수도권 주택시장에서 매매와 전세 가격 간의 시공간적 상호작용에 대한 국지적 모델을 구축하였다. 이를 통해 강남구의 아파트 가격의 대체효과를 살펴본 결과, 전역적 모델과 달리 강남구에서는 아파트 매매가와 전세가 간 대체효과가 발생하지 않았으며, 전세가격이 다른 재화에 독립적으로 존재하는 것으로 나타났다. 대신 강남

구 아파트 전세가는 상급재화로서 서울 외곽지역이나 근교의 연립다세대 혹은 단독다가구 주택의 전세가에 영향을 주었다. 외부적 호재에 의해 강남구 아파트 전세가격이 상승하게 될 경우, 아파트 가격은 각 지역별로 다른 시차를 두고 경기 외곽까지 영향을 받지만 조정 작용에 의해 4개월 이내에 대부분 소멸되었다. 또한 연립주택의 경우 비교적 가격이 저렴한 전세는 접근성이 좋은 주변지역으로 대체효과가 빠르게 나타났지만, 매매는 인천 지역 등 비교적 원거리에서 나타났다.

마지막으로 도출된 SpVAR-Lasso에 의한 모델을 기존의 전역적 SpVAR 모델과 비교함으로써 주택 가격 연구에서 새로운 방법론의 적용 가능성을 평가하였다. 추정과 예측의 오차를 측정한 결과, SpVAR-Lasso에서 정확성이 개선되었으며 실제 관측값과 매우 유사한 정도의 추정치와 예측치를 제시하였다. 또한 모델의 시공간적 효과의 필터링 여부를 살펴본 결과, SpVAR-Lasso 모델에서는 오차항에서의 공간적 자기상관성 문제가 다소 해결된 것으로 나타났다. 이에 따라 SpVAR-Lasso 모델은 기존 모델에 비해 정확하며 공간적으로 세밀한 해석이 가능하다는 점에서 주택 가격의 시공간적 상호작용 패턴 분석에 있어 충분히 활용될 수 있는 방법으로 볼 수 있다.

본 연구는 세분화된 주택 가격 간 시공간적 상호작용을 개선된 방법론을 통해 국지적으로 살펴볼 수 있었다는 점에서 의의가 있다. 기존의 방법론은 다수의 변수를 다루기에 부적절하여 다양한 주택 시장 가격 간의 관계를 포착하기 어려워 이에 관한 연구가 부족했기 때문이다. 이에 따라 본 연구에서는 Lasso를 통해 이러한 한계를 극복하고 보다 정확한 모델을 제시함으로써 수도권 주택의 시공간적 대체효과를 국지적 수준에서 실증하였다는 점에서 기존 연구와 다른 차별성을 갖는다.

주요어: SpVAR, Lasso, SpVAR-Lasso, 공간적 의존성, 공간적 이질성, 시공간 상호작용, 주택 매매-전세 가격, 수도권 주택 시장

학 번: 2014-20248

| 목 차 |

제 1 장 서 론	1
제 1 절 연구배경 및 연구목적	1
제 2 절 연구대상 및 연구방법	5
1. 연구대상	5
2. 연구방법	8
제 3 절 논문의 구성	9
 제 2 장 주택 가격 연구의 계량적 방법론	11
제 1 절 주택 가격 특성 연구	11
1. 주택 가격의 추정 및 예측	11
2. 주택 매매-전세가격 간 상호작용	14
제 2 절 주택 가격에 대한 계량 모델	18
1. 헤도닉 모델의 활용 및 한계	19
2. 시계열 모델의 활용 및 한계	24
 제 3 장 SpVAR-Lasso 모델링	31
제 1 절 주택 가격의 시공간적 특성 탐색	32
1. 주택 가격의 시계열적 정상성	32
2. 주택 가격의 공간적 의존성 및 이질성	35
제 2 절 주택 매매-전세가격의 시공간 상호작용 모델	40
1. SpVAR-SUR: 전역적 모델	41
2. SpVAR-Lasso: 국지적 모델	45
3. 모델의 비교 및 평가	52

제 4 장 SpVAR-Lasso 모델의 적용: 수도권 주택 가격	53
제 1 절 수도권 주택 가격의 시공간적 특성	53
1. 수도권 주택 가격의 시간적 특성 평가	53
2. 수도권 주택 가격의 공간적 효과 측정	61
제 2 절 수도권 주택 가격의 시공간적 상호작용 분석	67
1. 주택 매매-전세가격의 전역적 상호작용 패턴	67
2. 주택 매매-전세가격의 시공간적 대체효과	71
3. 주택가격의 시공간 상호작용에 대한 SpVAR-Lasso 적용 가능성	86
제 5 장 결 론	92
참 고 문 헌	96

| 표 목차 |

<표 1-1> 변수 및 변수 약어	7
<표 2-2> VAR모델의 분산분해 수식 기호(notations)	27
<표 3-1> SpVAR의 수식 기호(notations)	41
<표 4-1> 수도권 주택 종류별 거래가 평균 및 표준편차	54
<표 4-2> 2011년 1월 가격 기준 상위 5개 지역(만원/㎡)	56
<표 4-3> 2015년 6월 가격 기준 상위 5개 지역(만원/㎡)	56
<표 4-4> 평균 가격 기준 상위 5개 지역(만원/㎡)	56
<표 4-5> 가격 변화율 상위 5개 지역(%)	57
<표 4-6> 변수별 ADF 검정 결과(비정상 시계열 개수)	60
<표 4-7> 변수별 패널 단위근 검정 결과(t-bar statistic)	60
<표 4-8> 변수별 공간적 자기상관지수(Moran's I)	62
<표 4-9> SpVAR 추정 계수(SUR)	69
<표 4-10> 강남구 아파트 전세가 변화량에 대한 충격반응량(만원/㎡)	75
<표 4-11> 지역별 평균 추정 오차	88
<표 4-12> 예측오차의 공간적 자기상관지수(Moran's I)	91

| 그림 목차 |

[그림 1-1] 연구흐름도	10
[그림 3-1] LISA에 의한 지역 구분 모식도	38
[그림 3-2] 통상최소자승법과 능형회귀의 평균잔차제곱	48
[그림 3-3] 추정 회귀계수의 기하학적 이해	48
[그림 4-1] 수도권 주택 종류별 평균 거래가	55
[그림 4-2] 수도권 주택 가격 변화율	58
[그림 4-3] 가격변화량의 공간적 군집	64
[그림 4-4] GWR 회귀계수	66
[그림 4-5] 전역적 SpVAR 모델의 추정 결과	70
[그림 4-6] 강남구 아파트 전세가에 의해 영향을 받는 지역	72
[그림 4-7] 강남구 아파트 매매가에 영향을 주는 지역	74
[그림 4-8] 강남구 아파트 매매가에 의해 영향을 받는 지역	74
[그림 4-9] 강남구 아파트 전세가 변화에 의한 지역별 충격반응량	76
[그림 4-10] 강남구 아파트 전세가 변화량에 대한 수도권 아파트 전세가의 충격반응량	79
[그림 4-11] 강남구 아파트 전세가 변화량에 대한 수도권 아파트 매매가의 충격반응량	80
[그림 4-12] 강남구 아파트 전세가 변화량에 대한 수도권 연립다세대 전세가의 충격반응량	81
[그림 4-13] 강남구 아파트 전세가 변화량에 대한 수도권 연립다세대 매매가의 충격반응량	82
[그림 4-14] 지역별 아파트 전세가 변화에 대한 설명 비율	85
[그림 4-15] 지역별 평균절대오차(MAE)의 공간적 분포	90

제 1 장 서 론

제 1 절 연구배경 및 연구목적

2008년 글로벌 금융위기 이후 시작된 전세가의 상승은 2015년 현재까지도 지속되고 있는 반면 같은 기간 매매가는 상대적으로 정체되었다. 특히 2012년 6월부터 2013년 2월까지 주택 매매가격은 하락세를 보였지만 동기간 전세가격은 오히려 꾸준히 상승하였으며, 수도권외의 매매가 대비 전세가 비율은 2012년 1월 54.1%에서 2015년 6월 64.3%로 약 10%포인트 이상 높아졌다. 우리나라에만 있는 독특한 형태의 임대제도인 전세는 주거 형태의 약 20%(국토교통부, 2014)를 차지하는 주요 임차방식으로서 세입자들의 주거 안정성과 관련 깊다. 이에 따라 정부에서는 지나친 전세가 상승으로 인한 서민 피해를 줄이고자 일련의 전월세 대책(4.1 대책, 9.1 대책, 8.28 대책, 2.26 대책)을 시행하였으나, 그 효과가 미미한 편이다.

전세가의 급등이 오늘날에만 국한되는 현상은 아니지만 최근의 변화는 기존과 다른 양상을 보인다. 2001년 전세가격이 1년 동안 16.4% 상승하여 매매가 대비 전세가 비율이 64.6%에 달했을 때는 전세수요가 매매수요로 전환되면서 안정을 되찾을 수 있었다(최현일·임병진, 2011). 하지만 최근 금융위기 이후 '부동산 불패신화'가 지속되던 우리나라에서 주택가격이 오히려 떨어지는 현상이 발생하고 매매가격의 정체에 대한 우려 또한 지속되면서 전세가의 상승을 매매수요로의 전환을 통해 더 이상 해결할 수 없는 상황에 이르렀다(윤종인, 2015). 특히 과거 아파트를 중심으로 한 전세가의 상승이 이제는 단독주택이나 연립다세대 주택 등 다른 유형의 주택에서도 발견되는 현상이 나타나고, 매매가에 육박하는 전세까지 등장하면서 상황은 보다 복잡하게 전개되고 있다.

이러한 주택시장의 변화는 과거의 주택시장과 공간적으로도 다른 양상을 보이고 있다. 수도권외의 주택 매매가격이 보합 또는 하향안정세를 보이는 반면, 비수도권에서는 상승하는 추세가 나타난다. 전세의 경우 전국에

서 상승하는 패턴을 보이거나 특히 수도권에서의 급격한 상승세가 눈에 띈다(최두열, 2012). 이러한 주택시장의 공간적 변화 형태는 마치 유행이나 질병처럼 특정 지역에서 다른 지역으로 전이되는 현상이 있는 것으로 인식될 수 있다(김대원·유정석, 2015). 즉, 다른 지역의 지가 변화에 1분기 가량 선행하는 것으로 알려진 서울(김의준 외, 2000)은 주택시장 변화의 진원지로 간주할 수 있으며, 그 중에서도 특히 강남 지역의 아파트 가격 변화는 우리나라 부동산 시장을 평가하는 주요 지표로서 각종 정책의 대상 지역이 되고 있다(김경민, 2007). 하지만 국내 주택 시장에서 특정 지역의 주택 가격 변화가 다른 지역의 가격 변화에 영향을 주는지에 대해서는 여전히 논쟁적이며(이용만·이상한, 2004; 김경민, 2007; 서승환, 2007; 김호철·최창규, 2008), 주택 가격의 공간적 전이 현상을 포착하고자 하는 연구가 꾸준히 진행되고 있다(최두열, 2012; 김대원·유정석, 2015).

그러나 이들의 연구들은 지역별 또는 주택 유형별로 분절된 주택시장을 단일 주택시장으로 가정하거나 하나의 범주로 구분된 시장 내에서의 매매가와 전세가의 관계만을 분석했다는 점에서 한계를 지니고 있다. 최근 주택 시장의 변화를 살펴보면, 과거와 달리 아파트뿐만 아니라 다른 주택 유형에서의 전세가 역시 상당한 폭으로 변하고 있어 주택 유형별 가격 간에 발생하는 상호작용을 파악할 필요가 있다. 또한 최근 주택 가격의 변화가 공간적으로 이질적이며, 공간적 상호작용이 발생함에 따라 지역별로 구분되는 시장 가격 간의 관계 역시 새롭게 조명되어야 한다. 나아가 지역과 주택 유형이라는 두 범주의 주택 시장을 동시에 살펴봄으로써 다양한 주택 가격 간의 시공간적 상호작용 관계를 보다 심층적으로 살펴야 한다. 이러한 분석의 틀은 근래 발생한 주택 시장의 변화를 폭넓게 이해할 수 있게 도와주며 보다 효과적인 주택시장 안정화 정책을 수립할 수 있는 기초적 자료를 제시해준다는 점에서 의의를 갖고 있으나, 기존의 분석 방법들은 이를 효과적으로 다루기에 다소 미흡한 편이다.

주택 가격을 추정·예측하고 주택 가격(매매와 전세)간 상호작용을 분석한 다수의 연구에서는 다양한 실증적 분석 방법들이 적용되었다. 미시적 관점에서 주택의 속성 변수로 매매가격 또는 전세가격을 추정하는 헤도닉

모델들이 제안되었으며, 거시적 관점에서는 다양한 시계열 분석 방법을 활용하여 거시적 변수와 주택 가격간의 장기적 균형을 도출하는 연구가 진행되었다. 장기적 균형 관계를 입증하는 거시적 분석의 한 측면에서 매매가와 전세가의 상호 작용에 관한 연구들도 활발히 수행되었다. 이러한 연구들에 많이 활용된 벡터자기회귀모델(vector autoregressive model: VAR)은 자료의 시계열적 상관관계를 파악할 수 있다는 점에서 시간적으로 누적된 주택 가격 간 관계를 분석하는데 용이하였지만, 주택이 갖고 있는 공간적인 특성을 반영하지 못하였다는 점에서 한계를 보였다(Kueth and Pede, 2011). Beensotck and Felsenstein(2007)이 제시한 공간적 벡터자기회귀모델(spatial vector autoregressive model: SpVAR)은 VAR 모델이 갖는 공간적 한계를 극복한 방법론으로서 주택 시장에서 발생하는 변수 간의 시공간적 상호작용을 동시에 다룰 수 있다. 하지만 자료의 시계열 길이에 따라 모델에 추가될 수 있는 변수의 개수가 한정됨에 따라 세분화된 주택 시장에서 발생하는 매매가와 전세가의 상호작용 파악에 부족한 부분이 있다.

이에 따라 본 연구에서는 주택의 매매가와 전세가 간의 시공간적 상호작용 패턴을 파악하기 위해 기존의 방법론이 갖는 한계를 극복한 새로운 방법론을 제시하고, 이를 실제 수도권 주택 시장에 적용해봄으로써 새로운 모델의 활용 가능성을 탐구하고자 한다. 상기의 연구 목적을 달성하기 위한 세부 연구 주제는 세 가지로 구성된다. 첫 번째는 주택시장을 지역과 주택 유형, 거래 유형(매매, 전세)에 따라 구분하고 이들의 시공간적 특징을 파악함으로써 이들 간의 관계를 분석하기에 적합한 모델을 찾고자 한다. 두 번째는 시공간적 상호작용을 파악할 수 있는 SpVAR 모델의 추정 방식을 개선하는 것으로, 기존 모델의 통계적 문제를 해결하고 국지적 수준에서 주택 가격 간 관계를 살펴볼 수 있는 방법을 제안하고자 한다. 마지막으로 개량된 SpVAR 모델을 실제 수도권 지역에 적용함으로써 수도권 주택 시장에서 발생하는 매매와 전세 간의 시공간적 상호작용 패턴을 분석하고 새롭게 제시하는 모델의 적용가능성을 모색하고자 한다.

첫 번째 세부 연구목적인 주택 가격의 시공간적 특성 탐색은 주택 매매

-전세가격 간 상호작용에 적합한 모형을 찾는 사전적 분석으로서 필요하다. 주택 가격이 갖는 시간적 효과와 공간적 효과를 모두 살펴봄으로써 주택 가격 간 관계가 시공간 모델인 SpVAR에 적합한지 판단해야 하기 때문이다. 만약 주택 가격에 공간적 효과가 없는 것으로 나타난다면 복잡한 SpVAR 모델보다 간단한 VAR 모델이 쉽고 효과적으로 분석이 가능하며, 반대로 시간적 효과가 나타나지 않을 경우 공간자기회귀모형(spatial autoregressive model: SAR)이 보다 적합할 것이기 때문이다.

다음으로 기존의 SpVAR 모델이 갖는 통계적 문제를 해결함으로써 보다 정확하며 효율적인 모델을 제시하고자 한다. SpVAR 모델은 짧은 시계열에 비해 많은 변수를 다루고 있어 자유도 문제로부터 자유롭지 않았다. 기존의 추정방법은 이를 해결하고자 자료를 변수별로 통합했기 때문에 정보의 손실이 발생할 수밖에 없었으며, 이를 위한 공간적 등질성 가정이 전제되었다. 이러한 접근은 지역에 따라 다르게 설정되는 변수 간의 관계를 온전히 포착하기 어려워 추정이나 예측에 있어 그 정확도가 낮을 수밖에 없었으며, 오차항에서의 공간 관계를 추가적으로 정의해야 하는 과정상의 복잡성이 존재한다. 이에 따라 본 연구에서는 자료를 통합하지 않고도 효과적으로 추정할 수 있는 방법인 Lasso(least absolute shrinkage and selection operator)를 결합함으로써 정보의 손실을 방지하고 비현실적인 등질성 가정으로부터 자유로운 모델을 제안하고자 한다.

마지막으로 새로운 방법론(SpVAR-Lasso)의 적용 가능성을 평가하고자 실제 데이터를 활용하여 실증 분석을 실시한다. SpVAR-Lasso 모델을 통해 분석한 수도권 주택의 매매가와 전세가의 시공간적 상호작용 패턴을 통해 수도권 주택시장의 시공간적 역동성을 파악함으로써 모델의 활용성을 모색한다. 예를 들어, 특정 지역에서 특정 유형의 주택 전세가격 변화가 다른 주택 가격에 어떠한 영향을 주는지 파악하고자 한다. 또한 해당 지역의 주택 가격에서 발생한 변화가 다른 시장의 변화로 얼마나 설명되는지를 확인함으로써 수도권 주택 시장에서 발생하는 시공간적 상호 작용 패턴을 이해하고자 한다. 또한 기존의 방법과의 정확성 비교를 통해 모델의 개선 정도를 측정함으로써 새로운 방법의 유용성을 확인할 수 있을 것이다.

제 2 절 연구대상 및 연구방법

1) 연구대상

본 연구에서는 국내 주택 시장의 구조적 특성과 공간적 속성을 모두 반영하여, 시군구별 주택 유형에 따라 주택 하위 시장을 구성하고 이를 분석의 대상으로 설정하였다. 즉, 주택 시장을 소유형태별, 지역별, 유형별 하위시장(submarket)으로 구획하여 분석에 활용하였다. 주택의 하위시장이란 모호하지만 일반적으로 동일 하위시장 내에서 상호 대체가 가능한 주택들의 집합으로 정의된다(Bourassa *et al.*, 1999). 상호 대체성에 주목하는 하위시장 구획은 달리 말하면 동일 하위시장에서의 동질성이 확보되어야 함을 의미한다. 결국 한 하위시장 내에서는 주택의 구조, 입지, 지역적 특성이 무차별적이어야만 한다(Watkins, 2001). 하나의 등질지역이라는 가정 하에서 세워진 주택 가격 모델의 논리적 타당성은 물론 추정 및 예측 결과의 정확성을 높일 수 있기 때문이다.

이러한 측면에서 주택의 하위 시장을 구획하는 기준은 다양할 수 있지만 대표적으로 주택의 구조적 특성과 공간적 특성이 있다. 우선 구조적 특성은 주택이 갖는 건물의 특성을 의미하는 것으로, 방의 개수(Schnare and Struyk, 1976)나 집의 크기(Bajic, 1985; 임재만, 2013) 등이 있다. 주택의 유형도 하나의 기준으로서 미국의 경우, condominium, apartment, single family home으로 구분하였으며(Allen *et al.*, 1995), 국내도 유사하게 아파트, 연립다세대주택, 단독주택으로 구분하고 있다(임규채·기석도, 2006; 신종협·서대교, 2014). 이는 집합적 주거 양식과 거주 인구 밀도에 의해 주택 소비자들의 선호가 구분되고 있음을 보여주고 있다.

한편 주택의 공간적 특성은 주택 시장을 공간적 하위시장으로 구획하는 근거를 제공한다. 공간적 하위시장은 상호 대체성이 높은 주택들의 집합이 공간적으로 근접하여 존재하는 시장을 의미한다(손철, 2011). 이러한 공간적인 구획은 주택이 갖는 입지적 특성과 함께, 지역적인 속성을 공유한다는 점에서 근거를 찾을 수 있다. 이에 따라 동일한 행정 서비스를 제공받

고, 통계적 속성을 공유하는 행정구역 단위로 하위시장을 구획하거나 (Sonstelie and Portney, 1980; 김경민, 2007; 김대원·유정석, 2015), 센서스 등의 지역적 특성을 통계적으로 합역하여 등질지역을 구성하기도 한다(Ball and Kirwan, 1977; Watkins, 2001; 원제무 외, 2009; 최두열, 2012). 또한 이질적이라고 판단되는 선험적인 지식에 근거하여 공간 구획을 하는 경우도 있는데, 이 때 산이나 강 등의 자연지형을 기준으로 한다(Munro, 1986). 우리나라의 주택 문제에서 가장 큰 이슈가 되는 강남의 경우, 행정구역의 이름임과 동시 한강 이남을 지칭하는 일반 명사로서 그 구획이 개념적으로 뚜렷하지는 않지만, 다수의 인식 속에 동질성을 갖는 지역으로 인식되므로 ‘강남 3구’등 강남지역에 대한 조작적 정의를 통해 구획하여 분석의 대상으로 삼고 있다(이용만·이상한, 2004; 박동국·천인호, 2006; 김호철·최창규, 2008).

이상의 연구들을 통해 주택 하위시장의 구획 기준에 다양한 변수들이 활용될 수 있으나 여전히 명확한 준거가 없음을 확인할 수 있었다. 이는 결국 주택이 갖는 이질성을 어느 수준에서 동질적인 것으로 용인하고 전제로 하는가에 대한 문제와 연결된다. 이에 따라 하부 시장의 구획은 연구의 맥락과 모델의 사용 목적에 따라 달라질 수밖에 없으며, 이를 통해 보다 유용한 정보를 제공할 수 있는 방향으로 설정되어야 한다(Bourassa *et al.*, 2003).

이러한 맥락에서 수도권 주택시장을 주택의 구조적 특성으로 구분할 경우, 주택 유형이 적합하다. 수도권의 주택시장에서 아파트는 가장 큰 비중을 차지하는 주택 유형으로 다른 주택 유형에 비해 고가이면서 인구밀도가 높다. 반면에 연립주택과 단독주택의 경우 비교적 저가이며 낮은 밀도를 보이거나, 법률적으로 구분되기 때문에 이를 다른 주택 시장으로 보고자 한다. 또한 공간적 속성으로서 시군구의 행정단위는 다수의 연구를 통해 등질지역으로 다루어지고 있으며(김경민, 2007), 부동산 과열지구 설정 등의 부동산 정책이 시군구 단위를 대상으로 한다는 점에서 본 연구의 분석단위로 적합하다고 판단된다.

이러한 시군구별 주택 유형별 매매가와 전세가는 국토교통부가 제공하

는 월별 실거래가 자료를 활용하여 시군구 단위로 집계하였으며, 시간적 단위는 자료 제공 단위인 월을 그대로 사용하였다. 월별 실거래가 자료는 다른 부동산 관련 지수들에 비해 시간적 범위가 짧다는 한계가 있으나, 공간적 단위의 선택에 있어 자유로우며 주택 종류와 거래 형태별로 지수가 아닌 가격 형태로 제공이 되어 보다 현실적인 해석이 가능하다(임재만, 2004). 또한 주택 거래 당사자가 직접 신고한 자료를 토대로 이상치(outlier)들이 관리자의 선택적인 근거에 의해 제거된 형태로 공개되고 있어 신뢰성도 높은 편이라고 볼 수 있다. 분석의 대상이 되는 월별 실거래가는 주택 거래 신고가를 연면적(단독/다가구) 또는 전용면적(아파트와 연립다세대)으로 나눈 단위 가격(만원/m²)을 적용하여, 면적의 효과를 배제하였다. 이를 변수화하여 아래 <표 1-1>과 같이 변수 이름을 지정하였다.

<표 1-1> 변수 및 변수 약어

	아파트(A)	연립다세대(B)	단독다가구(C)
전세(Rent)	RA	RB	RC
매매(Price)	PA	PB	PC

분석의 공간적인 범위는 서울특별시, 인천광역시, 경기도를 포함하는 수도권 65개 시군구이다. 인구의 49%가 집중된 수도권에서 주택 가격 변화는 타 지역에 비해 빠르고 급격하며, 최근에 촉발된 전세가 상승 역시 수도권에서 크게 나타나고 있다. 이러한 수도권의 주택 가격 변화는 전세난 등의 사회적 이슈를 촉발하기 때문에 그 변화의 양상을 면밀히 살펴볼 필요가 있다. 이 때 인천광역시 옹진군의 경우 아파트와 같은 특정 형태의 주택이 존재하지 않아 분석의 대상에서 제외하였다. 또한 시간적 범위는 2011년 1월부터 2015년 6월까지의 총 54개월이며, 2015년 7월 자료는 모델의 예측력 평가 자료로 활용하였다.

2) 연구방법

연구 방법은 전술한 세 가지 세부 연구 목적을 달성하기 위해 크게 탐색적 분석 - 모델링 - 모델의 적용 및 평가의 3단계를 거친다. 탐색적 분석 과정에서는 주택 가격의 시공간적 특성인 시계열적 정상성(temporal stationary)과 공간적 의존성(spatial dependancy), 공간적 이질성(spatial heterogeneity)을 중심으로 평가하고자 한다. 시계열적 정상성은 수정보완된 딕키-풀러 검정(augmented Dickey-Fuller: ADF)과 패널 단위근 검정(panel unit root test)을 통해 확인하며, 공간적 의존성과 이질성은 공간적 자기상관지수와 지리가중회귀(geographically weighted regression: GWR)를 통해 살펴본다.

모델의 구축 단계에서는 차원 회귀분석에서 변수 선택에 효과적인 Lasso를 SpVAR 모델의 추정에 접목하고자 한다. Lasso 추정은 관측사례 수보다 변수의 개수가 많은 고차원 자료에서 효과적으로 설명변수를 선별할 수 있는 방법으로서 다양한 분야에서 활용되고 있다. 이는 복잡한 시공간 자료를 동시에 다루는 SpVAR 모델에서 발생하는 자유도 문제를 다룸에 있어 매우 효과적이며, 다수의 주택 가격 간의 시공간적 상호작용을 파악하기 위해 적합한 방법으로 활용될 수 있다. 그러므로 본 연구에서는 SpVAR 모델을 Lasso로 추정하는 SpVAR-Lasso 모델을 제안하며, 이를 바탕으로 기존 VAR 모형에서 계수의 의미를 해석하고자 활용된 충격반응 분석과 분산분해를 공간적으로 확장하여 시도할 것이다.

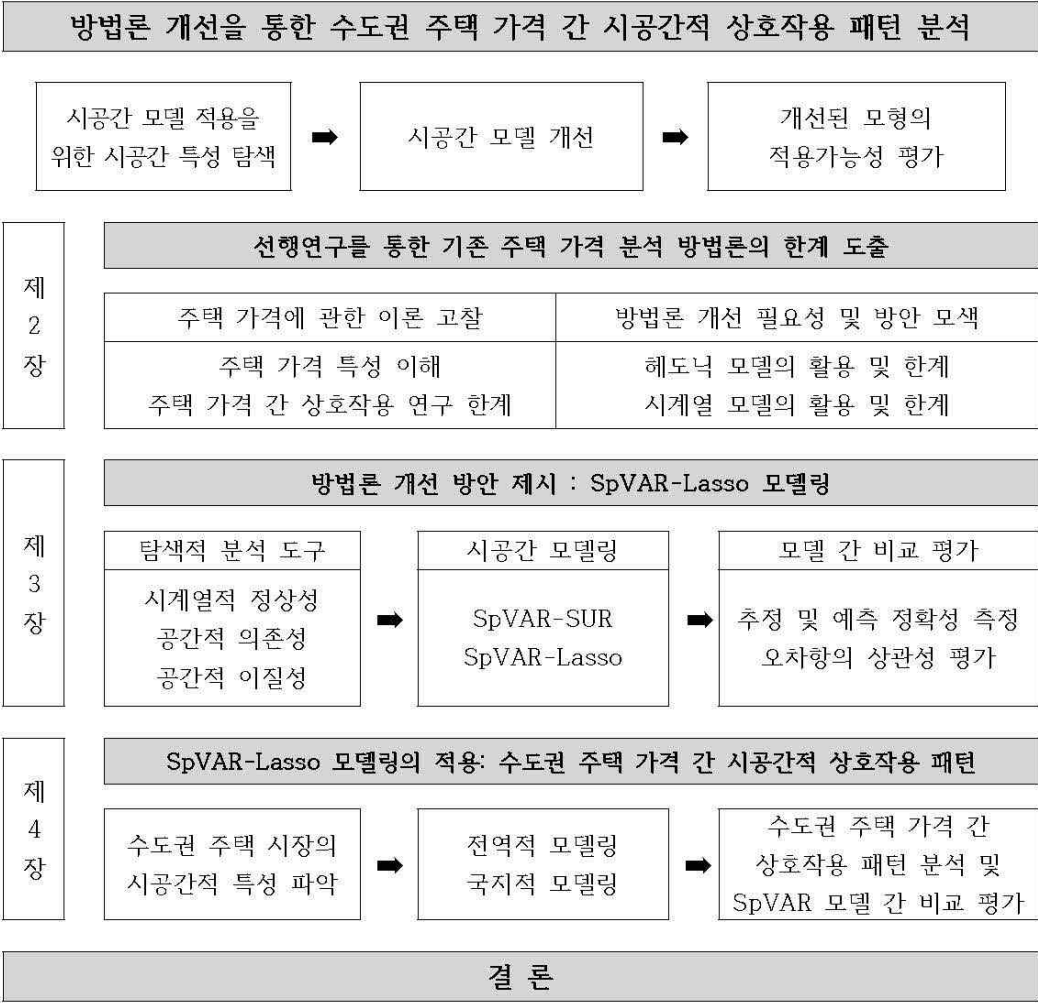
마지막으로 SpVAR-Lasso 모델을 수도권 주택 시장에 적용하고 그 활용 가능성을 평가하고자 한다. 이를 통해 실제 주택 시장에서 나타나는 주택의 매매가와 전세가 간의 시공간적 상호작용을 파악하고 기존의 방법론으로 보기 어려웠던 패턴을 포착함으로써 새로운 모델의 유용성을 평가하고자 한다. 또한 새로운 방법론과 기존 방법론의 추정과 예측 정확성을 비교하고 오차에서의 공간적 자기상관 유무를 측정함으로써 모델의 정밀성을 검토하고자 한다. 이 때, 추정과 예측 결과의 정확성은 평균절대오차(mean absolute error: MAE)와 평균제곱근오차(root mean square er-

ror: RMSE)를 통해 비교하며, 오차의 공간적 자기상관성 유무는 Moran's I 지수를 통해 확인한다.

일련의 분석과정은 R 3.1.2를 통해 실시하였으며, *plm*, *spdep*, *vars*, *glmnet* 등의 패키지를 활용하였다. 또한 분석 결과는 ArcGIS 10.1를 사용하여 지도형태로 시각화함으로써 공간적 변이를 보다 선명하게 제시하고자 하였다.

제 3 절 논문의 구성

본 연구의 구성은 다음의 [그림 1-1]과 같다. 우선 2장에서는 선행연구들 통해 주택 가격의 특성과 매매가격과 전세가격 간 상호작용에 관한 이론을 이해하며, 이를 분석하고자 활용된 방법론들의 한계를 살펴봄으로써 방법론 개선의 필요성을 검토한다. 3장에서는 본 연구가 제안하는 SpVA-Lasso 모델링을 제시하고 이를 위한 탐색적 분석, 모델의 활용, 모델의 평가를 위한 일련의 방법들을 설명하고자 한다. 구체적으로 자료의 시공간적 특성이 SpVAR-Lasso 모델에 적합한지 확인하고, 이를 모델에 적용한 뒤 정확성이 개선되는지를 확인한다. 4장에서는 3장에서 다룬 방법론을 실제 수도권 주택 시장에 적용하여 수도권 주택 매매가와 전세가 간 시공간적 상호 작용 패턴에 대한 파악하고 이를 통해 SpVAR-Lasso 모델의 활용 가능성을 평가하고자 한다. 마지막으로 5장에서는 분석 결과를 요약하고 정리하여 본 연구의 의의와 한계를 살펴보고자 한다.



[그림 1-1] 연구흐름도

제 2 장 주택 가격 연구의 계량적 방법론

본 장에서는 선행연구를 통해 사례 분석의 대상이 되는 주택 가격의 일반적인 특성을 고찰하고 이를 반영한 계량적 방법론을 조망함으로써 기존에 활용된 방법론들의 성과와 한계를 파악하고자 한다. 이는 본 연구가 제안하는 새로운 방법론의 필요성을 확인하고 이를 주택 가격 간 상호작용 패턴 분석에 활용할 수 있는지를 검토할 수 있는 이론적 근거를 제시하기 위함이다. 보다 구체적으로 1절에서는 주택 가격의 추정 및 예측과 매매-전세 간의 상호작용에 관한 이론을 살펴봄으로써 주택 가격이 갖는 고유한 특성들을 확인하고 기존의 연구들이 간과한 주택 시장에서 발생할 수 있는 공간적 상호작용을 파악하고자 한다. 2절에서는 1절에서 밝힌 주택의 시공간적인 특성을 반영하기 위한 방법론들을 살펴본다. 이를 통해 시공간적 효과를 모두 갖고 있는 주택 가격 간 상호작용을 보기 위한 시공간 통합 모델의 필요성을 역설하고, 통합 모델 활용 시 발생할 수 있는 문제를 제시함으로써 본 연구의 당위성을 확보하고자 한다.

제 1 절 주택 가격 특성 연구

1) 주택 가격의 추정 및 예측

주택은 주거서비스를 제공하는 소비재임과 동시에 투자 가치가 있는 자산의 성격을 모두 갖고 있어 다른 재화들과 구분되는 시장을 형성하며, 시장에서 형성되는 가격 역시 일반적인 재화의 가격 형성 과정과 달리 복잡하여 이에 대한 다양한 이론이 존재한다. 주택 가격이 형성되는 과정을 보는 관점은 크게 미시적 관점과 거시적 관점으로 나뉜다. 주택의 공급이 고정되어 있거나 매우 비탄력적이라고 보는 미시적 관점에서 주택 가격은 주택이 제공할 수 있는 서비스의 총합으로 본다. 즉, 주택 가격이란 주택이 갖는 다양한 특성들의 가치와 양으로 설명된다는 입장이다. 헤도닉 모

델(hedonic price model)은 이러한 미시적 관점의 주택 가격 형성 프로세스를 이해하는 주요한 방법으로 주택 가격의 추정에 널리 활용되고 있다(Lee *et al.*, 2015).

헤도닉 모델은 Rosen(1974)에 의해 균형 시장 가격을 도출할 수 있는 방법으로 증명된 이후 주택 가격 추정에 폭넓게 활용되었다. 기본적으로 재화의 가격을 특성들의 양에 대해 회귀함으로써 그 가격을 추정하는 헤도닉 모델은 주택이 갖고 있는 구조적인 특성, 입지적 특성, 근린 지역의 특성 등을 반영한다. 초기 헤도닉 모델에서는 변수들 간의 선형적 관계에 초점을 맞추어 주택 가격과 속성 간의 관계를 살펴보았다. Harrison and Rubinfeld(1978)은 주택 가격과 깨끗한 공기와의 관계를 조망함으로써 주택 가격이 주변 환경에 영향을 받음을 증명하였다. 한편 Goodman(1978)은 주택이 갖는 속성을 구조적 변수, 근린 변수로 구분하여 오늘날 헤도닉 모델에서 변수의 범주를 구분하는 기준을 제시하였다. 이후 주택이 갖는 시공간적 특성을 보다 정교하게 반영한 모델을 적용함으로써 주택 가격이 주택의 고유한 구조적 특성뿐만 아니라 다수의 사회·경제적 요소들에 의해 설명될 수 있음을 밝혔다(Can, 1992; Pace *et al.*, 1998; Fotheringham *et al.*, 2002; Anselin and Lozano-Gracia, 2008; Huang *et al.*, 2010; Dubé and Legors, 2013; Parent and Hofe, 2013; Bárcena *et al.*, 2014).

국내에서도 이러한 헤도닉 모델을 통해 다양한 사회·경제적 변수들로 주택 가격을 추정하는 연구가 활발히 진행되었다. 이번송 외(2002)은 우리나라의 대표적 주택 유형인 아파트의 단지적 특성을 통하여 아파트 가격을 추정하였고, 강영옥(2004)은 고용지역, 수변공간과의 접근성을 고려하여 아파트 가격을 설명하고자 하였다. 뿐만 아니라 다수의 연구에서 주변 편의시설의 유무, 역세권, 교육시설, 교통 접근성, 심미적 요인 등 거주에 필요한 다양한 서비스를 통해 주택 가격을 설명하고자 하였다(박헌수, 2003; 이인화·문영기, 2007; 강창덕, 2010; 김성우·정건섭, 2010; 전경구, 2012; 김소연·김영호, 2013; 진찬우·이건학, 2014).

한편 거시적 관점에서는 개별 주택 가격이 아닌 주택 시장 전체의 동태

적 변화에 초점을 맞추어, 주택 가격을 결정하는 주택 시장 외부의 거시경제적 요인들과의 관계를 조망하고자 한다. 이는 주택 시장이 전체 거시 경제의 큰 비중을 차지하고 있어 다른 거시경제적 변수와 상호 밀접한 관련을 맺고 있기 때문이다(Leung, 2004). 이에 따라 다양한 거시경제 변수들은 주택에 대한 수요와 공급에 영향을 미치게 되는데, 이때 주택 가격은 수요와 공급의 장기적 균형을 달성하려는 조정 과정에서 형성된다(Chen and Patel, 1998). 즉 주택 가격은 주택의 수요와 공급을 형성하는 함수들로 표현할 수 있다(식 2-1).

$$\text{주택가격} = f(\text{인구, 실질소득, 통화량, 건설비용,...}) \quad [\text{식 2-1}]$$

이러한 거시적 관점의 연구에서는 주로 시계열 분석(time series analysis)을 활용하여 주택 가격에 영향을 주는 거시경제 변수를 탐색하고 이들 간의 상관관계를 밝힘으로써 미래의 주택 가격에 대한 예측을 시도하고자 하였다. Poterba *et al.*(1991)은 1980년대 미국 대도시의 주택 가격의 변화를 수요적인 측면에서 인구의 증감으로 설명하였으며, Abraham and Hendershott(1996)은 수요와 공급 측면에서 각각 실질소득과 실질건설비용이 주택 가격을 결정한다고 보았다. 또한 Chen and Patel(1998)은 타이페이의 주택 가격이 건설비용, 이자율, 총 가구실질소득, 주택 거래량, 주가와와의 장기 균형 관계에 있음을 밝혔으며, Kan *et al.*(2004)과 Davis and Heathcote(2005)는 지역에 따라 정도의 차이는 있지만 주택 가격과 국내총생산(GDP) 간의 밀접한 관련이 있음 실증적으로 규명하였다. 이외에도 다수의 연구들이 금리 또는 물가와 주택가격 간의 관계를 실증하고자 하였다(Anari and Kolari, 2002; Aoki *et al.*, 2002; Lastrapes, 2002; Tsatsaronis and Zhu, 2004; Giuliadori, 2005).

국내 주택 시장 역시 전체 자산 시장에서 차지하는 비중이 큼에 따라 거시적 관점의 주택 가격 연구가 활발히 진행되었다. 김경환 외(1991)의 연구에 따르면 물가와 주택 가격은 상호 의존적인 관계를 갖는 것으로 나타났다. 박헌수·유은영(2014)은 전국 16개 시도지역의 2006년부터 2012년

까지의 아파트 가격변화를 수요적 요인(소득, 인구, 조세, 금리)과 공급적 요인(토지가격, 주택건설비용, 정부규제) 간의 균형을 통해 설명하고자 하였다. 이 때, 주택거래량, 건설비용, 전세매매비율, 실질소득, 경제활동인구는 아파트 가격과 양의 관계를 갖는 반면, 이자율, 실업율, 고령화 비율은 주택가격을 하락시키는 요소로 나타났다. 금융적 요소와의 관계를 살핀 정규일(2006)의 연구에서는 주택가격과 유동성간의 장단기 관계를 실증적으로 분석하여 주택 가격은 주식과 달리 유동성에 장·단기적으로 모두 영향을 받음을 밝혔으며, 이영수(2008)는 금리가 상승하게 되면 19개월 후의 주택 가격이 2.8% 하락하는 결과를 도출하였다.

2) 주택 매매-전세가격 간 상호작용

한편 고가인 주택은 그 가치가 일시에 거래되거나 나누어 거래되곤 한다. 이를 주택의 매매와 임대로 구분할 수 있는데, 해외와 달리 우리나라에는 전세라는 독특한 임대제도가 존재한다. 전세는 매매와 월세의 중간적 형태로 독자적인 가격체계를 가지고 있으며(조주현·임정호, 2004), 주택 가격이 지속적으로 상승할 것으로 기대되는 경우 임대인과 임차인 모두에게 유리한 제도이다. 임대인은 주택을 구매할 때, 전세금을 통해 부족한 재원을 마련하며 이후 주택가격의 상승을 통해 이익을 추구하며, 임차인의 경우 전세금을 미래 주택 구입 준비 자금으로 활용함과 동시에 주거 서비스를 제공받기 때문이다(전해정, 2012). 결국 전세는 일정기간의 임대료 또는 그 이상을 일시에 지불하지만 원금은 유지된다는 점에서 독특한 성격을 지니고 있으나 근본적으로 주택이 제공하는 주거 서비스에 대한 사용비용, 즉 주택에서 창출되는 수익이라는 점에서 임대료의 일종이다.

이러한 전세의 가격은 주택이 갖는 고유한 특성으로 인해 결정되기도 하지만(김현재, 2003; 김태호·이창무, 2006), 매매가격과의 관계 속에서 가격이 형성된다. Gallin(2008)은 다음의 [식 2-2]를 통해 주택의 임대료와 매매가격 간의 관계를 정의하였다.

$$R_t = P_t [(i_t + \tau_t^p)(1 - \tau_t^y) + \delta_t + \Lambda_t - E_t G_{t+1}] \quad [\text{식 2-2}]$$

[식 2-2]에서 한 시점 t 에서의 임대료 R_t 는 주택의 현재가격 P_t 와 주택을 보유함에 따라 발생하는 사용자 비용(i_t : 이자율, τ_t^p ; 주택 보유세율, τ_t^y : 한계소득세율, δ_t : 유지비 및 감가상각율), 위험 프리미엄(Λ_t), 현재 기대되는 미래 예상 자본 차익 $E_t G_{t+1}$ 의 함수로 구성된다. 이 식을 통해 임대료는 사용자 비용이 높을수록 상승하며, 주택의 미래가치가 높을수록 하락함을 알 수 있다.

이러한 임대료(전세)와 매매가의 관계에 관한 연구는 크게 두 가지로 나뉜다. 첫 번째는 임대료와 매매가의 관계, 즉 두 변수 간의 비율 변화의 요인을 찾는 연구로서 임대료와 매매가에 영향을 주는 요소들의 변화를 파악하는 것을 목적으로 한다. Campbell *et al.*(2009)은 매매 대비 임대료 비율을 구성하는 요소인 사용자 비용(무위험 이자율), 주택 프리미엄, 임대료 상승 간의 관계를 규명함으로써 위험 프리미엄의 변화가 임대료 비율을 결정하는 주요한 요소이며, 세 요소 간의 공분산이 임대료 비율의 변화를 심하게 만듦을 증명하였다. 또한 Julliard and Wong(2008)은 OECD 국가들을 대상으로 분석한 결과 금융시장 발전지수와 지역별 규제 정책이 주택 프리미엄에 영향을 미침으로써 전세-매매 비율을 좌우한다고 주장하였다. 국내 시장에 관해서는 손재영(2000)이 자산 시장의 균형조건을 근거로 이자율과 가격상승이 전세-매매 가격 비율에 영향을 주고 있음을 밝혔고, 이용만(2000)은 주택 가격의 기대상승률 하락 혹은 주택 내재가치의 상승이 전세-매매 가격을 올리는 요소로 작용할 수 있음을 설명하였다.

한편 전세와 매매가의 관계 관한 두 번째 연구 주제는 임대료와 매매가 간의 상호작용에 관한 연구로서 이러한 연구들은 두 변수 간 선후행관계를 밝히고 그 영향력을 파악하고자 한다. 주택의 매매가와 전세가 간의 관계를 설명하는 가설로는 가격환원설과 제약선택설이 있다(임정호, 2006). 가격환원설이란 Ricardo의 지대론에서 출발한 오랜 가설로서, 주택 가격을 미래 내재가치의 가격환원부분으로 봄으로써 주택의 중심가격을 주거

서비스에 대한 사용료인 임대료로 인식하는 가설이다. 이러한 관점에서 임대가격은 독자적인 가격체계에 의해 결정되고 변동되며, 자산 시장의 균형을 맞추고자 매매가격이 변화한다(Phillips, 1988). 또 다른 가설인 제약선택설은 주택의 매매시장이 임대시장에 비해 우등재로서의 지위를 가지므로 매매가격의 독립성을 지지한다(Rosen, 1979). 이러한 관점에서 주택의 중심 가치는 매매가격과 임대가격의 차이에서 발생하는 자본 이득으로서 매매가격이 변할 경우에 자산 시장 균형을 위해 임대가격이 변한다고 보았다.

위에서 살펴본 두 가설에서 모두 전제하고 있는 자산시장의 균형 조건이란 주택을 통한 수익이 다른 대체투자(금융자산 등)에 의한 수익과 같음을 의미한다. 자산시장의 균형조건은 아래 [식 2-3]으로 표현할 수 있다.

$$r = \frac{R}{P} + \frac{\Delta P}{P} \quad [\text{식 2-3}]$$

시장이자율을 r , 임대료와 매매가격을 각각 R 과 P 로 나타낼 때, 시장의 이자율은 주택의 사용 가치와 자본이득의 합과 균형을 이룬다.

이러한 균형 속에서 시장 상황이 변함에 따라 임대료 혹은 가격이 변해야 하는데, 어느 것이 먼저 반응하고 다른 것을 변화시키는 요인으로 작용하는지는 여전히 논쟁적이다. 처음 이 문제를 다루게 된 계기였던 미국의 70년대 주택가격의 상승과 임대료의 하락에 대해 과연 어떠한 현상이 원인이었는지에 대한 논쟁이 치열하게 진행되었다. 임대 가격의 변화가 원인이라고 본 입장에서는 주택 소유주가 세후 주택 가격이 인플레이션으로 인해 떨어질 것으로 기대하여 임대료를 낮추거나(Titman, 1982), 사용 가치가 떨어질 것으로 예상하였다(Swan, 1984)는 점을 주목하였다. 반면 가격의 상승이 임대료의 변화를 이끌었다는 관점에서는 인플레이션에 대한 예측으로 인한 자가 소유 수요의 증가(Lowry *et al.*, 1984), 낮은 이자율로 인한 주택 구입 비용 감소(Feldstein, 1980) 등의 이유로 주택 매매 수요가 증가하여 주택가격이 상승하였다고 주장하였다.

국내에서는 전세와 매매가의 선후행에 관한 다수의 연구가 진행되었지

만, 두 거래 가격 중 어떤 것이 선도하는 요인인지를 설명하는 이론은 하나로 수렴되지 않은 상태이다(임재만, 2013). 우선 전세가에 의해 매매가가 결정된다고 본 연구들은 황두현(1990), 박동국·천인호(2006)의 연구를 꼽을 수 있다. 황두현(1990)은 1986년 1월부터 1990년 4월까지의 주택매매지수와 전세가격지수를 비교한 결과 전 지역과 주거형태에 관계없이 주거서비스에 대한 비용인 전세가격이 주택 매매가격을 형성해가고 있음을 밝혔다. 또한 박동국·천인호(2006)는 외환위기 이후의 주택 가격의 변동을 서울을 강남과 강북으로 나누어 살펴본 결과, 강남의 경우 전세가격이 매매가격을 설명하였다. 이는 강남지역 주택의 내재가치가 임대료에서 발생되고 있음을 지지하는 것으로 강남지역 주택은 자산으로서의 가치보다 실제 사용가치가 크게 나타나고 있음을 의미한다고 볼 수 있다.

한편 매매가격이 전세가격을 선도한다는 입장의 연구는 임재만(2004), 조주현·임정호(2004), 전해정(2012)등이 있다. 임재만(2004)은 그랜저 인과성 검정(Granger, 1969)을 통해 전세가격이 매매가격을 설명하지 못함을 밝혔고 나아가 매매가격이 1기와 2기 이전의 전세가격과 매매가격으로 구성되고 있다고 보았다. 이는 매매시장이 전세시장의 우등재로서의 역할을 하며 매매의 대안으로 전세를 택하고 있음을 반증하는 결과이다. 또한 조주현·임정호(2004)의 경우 역시 전세의 매매 대체가능성에 주목하였는데, 매매가격에 의한 충격은 전세시장에 크게 미치는 반면 역의 경우는 크지 않음을 보였다. 전해정(2012)은 글로벌 금융위기 이후 서울지역에서는 매매가가 전세가에 선행하며, 3~4기 전에 나타난 매매시장에서의 충격이 전세시장으로 퍼진다고 보았다.

그러나 이상의 연구들은 주택 시장이 갖는 이질성을 온전히 고려하지 못하고, 대부분이 단일 시장을 가정하거나 아파트와 같은 하나의 하위시장에서의 변화를 설명하였다. 그 결과 주택 시장 내부에서의 다양한 상호작용을 포착하지 못하고 다른 지역 또는 다른 하위시장과의 관계를 간과함에 따라 설명이 부족할 수밖에 없었다. 특히 주택 가격과 임대료는 공간적으로 주변 지역과 상호 작용을 하고 있음이 다수의 연구들(Giussani and Hadjimatheou, 1991; Stein, 1995; Pollakowski and Ray, 1997;

Mean, 1999; Beenstock and Felsenstein, 2007; Kuethé and Pedé, 2011; 이용만·이상한, 2004; 김경민, 2007; 서승환, 2007; 김대원·유정석, 2015)을 진행되기도 하였지만, 이들의 연구는 매매가 혹은 전세가만을 대상으로 할 뿐 이 둘 간의 공간적 상호작용은 포착하지 못하였다.

이에 따라 본 연구는 이견이 존재하는 전세와 매매 간의 상호작용에 관해 보다 세밀한 수준에서 살펴보고자 한다. 이미 임규채·기석도(2006)와 신종협·서대교(2014)는 주택의 종류별 전세와 매매가 간의 관계를 살펴보았으나 전국을 단일 시장으로 가정하고 분석하였다. 또한 가격의 공간적 확산과정을 살펴본 연구에서는 각각의 가격의 공간적 상호작용 패턴만을 확인하였으며, 구조적으로 이질적인 주택 유형을 간과하였다는 점에서 주택시장에서의 시공간적 상호작용을 온전히 포착하는 데 한계를 보인다. 그러므로 본 연구에서는 기존의 연구들이 간과한 주택의 이질성을 보다 현실적으로 반영하기 위하여 주택의 종류와 지역이라는 두 가지 기준을 통해 하위시장을 구획하고, 지역별 주택 유형에 따라 다르게 형성되는 전세와 매매 간에 발생하는 시공간적 상호작용에 대해서 살펴볼 것이다.

제 2 절 주택 가격에 대한 계량 모델

앞선 절에서는 주택의 가격의 추정과 예측을 통해 주택 가격에 나타나는 시공간적 특성을 살펴보았고, 주택 매매시장과 임대시장 간의 상호작용에 관한 이론적 근거를 조망하였다. 이를 통해 기존의 연구들이 갖는 한계를 도출할 수 있었는데, 이는 분석에 사용된 방법론의 한계에서 비롯된 것이라고 판단된다. 이에 따라 본 절에서는 앞서 살펴본 주택 가격의 속성과 현상을 분석하기 위한 방법을 집중적으로 살펴보고자 한다. 이는 각각의 방법론들이 갖는 타당성과 합목적성을 확인하기 위함이며, 나아가 각 연구들이 갖는 한계가 방법론의 한계에서 비롯된 것임을 살펴보고자 한다. 이러한 과정을 통해 본 연구에서 제시하고자 하는 SpVAR의 개선 방법의 논리적 타당성과 연구 목적과의 적합성을 파악할 수 있을 것이다.

1) 헤도닉 모델의 활용 및 한계

미시적 관점에서의 주택 가격을 추정하는 방법으로 활용되는 헤도닉 모델(hedonic model)은 기본적으로 다중회귀분석 방법을 사용한다. 다수의 독립변수들을 통해 하나의 종속변수를 설명하는 다중회귀분석은 주택이 갖는 구조적인 속성뿐만 아니라 입지적 특성, 지역적 특성 등으로 주택 가격이라는 하나의 변수를 추정하는데 적합한 방법이다.

$$y_i = \beta_0 + \sum_k \beta_k x_{ik} + \epsilon_i \quad [\text{식 2-4}]$$

[식 2-4]는 헤도닉 모델의 일반식으로, y_i 는 i 지역에서 관측된 주택 가격 또는 임대료를 의미하며, x_{ik} ($k=1, \dots, K$)는 상수항을 포함한 독립변수의 관측치, β_k 는 추정될 회귀계수, ϵ_i 는 랜덤 오차항이다. 이 때 x_{ik} 가 각각의 설명변수, 즉 주택을 구성하고 있는 속성들의 양이라면, 이에 상응하는 회귀계수 β_k 는 각각의 속성들이 해당 주택에서 차지하는 가치의 비율을 의미한다. x_{ik} 의 경우 주택을 조사하거나 주변의 관측을 통해 실제 값을 구할 수 있는 반면, β_k 는 추정을 통해 알아야 한다. 이를 위해 [식 2-4]를 행렬식 [식 2-5]로 표현할 경우, 회귀계수 벡터 β 는 통상최소자승법(ordinary least square: OLS)으로 불리는 [식 2-6]과 같이 추정할 수 있다.

$$y = X\beta + \epsilon \quad [\text{식 2-5}]$$

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad [\text{식 2-6}]$$

헤도닉 모델이 처음 주택 가격 추정에 활용될 당시 OLS와 같은 기초적인 방법이 주로 적용되었다. 앞서 살펴본 Harrison and Rubinfeld(1978)와 Goodman(1978)의 연구가 대표적인 사례이다. 이러한 방법은 분석의 용이성과 선형적인 지식에 의한 변수의 가격에 대한 영향력을 직관적으로

파악할 수 있다는 장점으로 국내 사례에도 활발히 활용되었다(이번송 외, 2002; 이인화·문영기, 2007).

그러나 통상최소자승법으로 회귀계수를 추정함에 있어 오차항에 기본적인 전제가 성립되어야 한다. 오차항의 분포가 독립적이며 동일해야 하는 가정(independent and identically distributed: i.i.d.)을 만족시켜야만 추정된 회귀계수를 신뢰할 수 있기 때문이다. 하지만 주택과 같이 공간상에 분포하는 자료는 공간적 효과(spatial effect)로 인해 이러한 가정을 만족시키기 어렵다. 공간적 효과란 어떠한 대상이 공간상에 위치함으로써 발생하는 특성으로서 공간적 의존성(spatial dependency)과 공간적 이질성(spatial heterogeneity)로 구성되어 있다(Anselin, 1988). 그 중 공간적 의존성은 어떠한 대상의 상대적인 위치로 인해 발생하는 효과로서 특정 영향력이 거리에 따라 감소되는 현상과 관련이 깊다. 공간적 의존성이 발생하는 원인으로 관측 단위의 임의성, 공간적 합역의 문제, 공간적 외부성(spatial externality)과 이전효과(spillover effect), 현상의 공간적 상호작용 등을 꼽을 수 있다(Anselin, 1988). 특히 주택은 가격에 영향을 미치는 근린 요소를 공유하면서 동시에 개발된다는 점에서 공간 단위의 문제가 발생하며, 주택 가격이 주변 지역에 미치는 외부효과 역시 발생된다. 게다가 실질적으로 주택 가격을 책정하는 부동산 업체들이 책정 가격의 기준을 주변 시세로 정하기 때문에 주택 가격은 공간적으로 의존적일 수밖에 없다(Fotheringham *et al.*, 2015a). 또한 공간적 이질성은 대상의 절대적인 위치에 의해 발생하는 것으로서 입지에 따라 특정 현상이나 현상 간의 관계가 다르게 나타나는 특성을 의미한다. 주택은 지역에 따라 선호하는 가치가 다를 수 있고, 비탄력적인 수요와 공급으로 인해 발생하는 주택 하위시장의 존재 여부(Goodman and Thibodeau, 2003)로 인해 주택 가격과 그것을 설명하는 속성들 간의 관계가 지역에 따라 다르게 나타난다(Fotheringham *et al.*, 2015a).

이러한 공간적 효과로 인해 주택 가격을 추정함에 있어 공간적 헤도닉 모델(spatial hedonic model)은 전통적인 헤도닉 모델이 갖는 통계적 한계를 극복할 뿐만 아니라 주택이 공간상에 입지함으로써 발생하는 속성을

보다 잘 반영할 수 있는 방법으로 널리 활용되었다. 대표적으로 Anselin(1988)이 제시한 공간자기회귀모델(spatial autoregressive model: SAR)과 Brunsdon *et al.*(1998)에 의해 제안된 지리가중회귀모델(geographically weighted regression: GWR)이 공간적 헤도닉 모델에 적용되었다. 우선 공간자기회귀모델을 살펴보면 다음의 [식 2-7]과 같다.

$$y = \rho Wy + X\beta + \epsilon \quad [\text{식 2-7}]$$

전통적인 헤도닉 모델인 [식 2-5]와 비교했을 때, ρWy 항이 우변에 추가되었음을 알 수 있다. ρWy 는 종속변수의 공간적인 자기상관향으로서 근린지역으로 정의된 범위(공간가중행렬, W)에서 발생하는 공간적 의존성을 의미한다. 이 때, Wy 는 공간적으로 지연된 공간차(spatial lag)이며, ρ 는 공간회귀계수를 의미한다. 이는 종속변수, 즉 주택 가격에 공간적 의존성이 발생하는 경우 적합한 방법이다. 그러나 주택 가격을 추정하는 문제에서 공간적 의존성은 비단 종속변수에서만 발생되지 않는다. 주택이 갖는 속성들 역시 공간적으로 영향을 주고받을 수 있으며, 관측되지 못한 변수들에서 공간적인 자기상관이 나타날 수도 있기 때문이다. 이러한 경우 오차항에서 공간적 자기상관이 측정되는데, 이를 반영하는 모델을 공간오차모델(spatial error model: SEM)이라고 한다(식 2-8). 이러한 공간적 의존성을 반영한 공간계량모델(spatial econometrics model)들을 활용한 연구는 Can(1992)의 연구를 필두로 오늘날까지 활발히 진행되며 주택 가격 추정에서 발생하는 공간적 의존성을 실증적으로 검증하고 있다(Dubé and Legors, 2013; Parent and Hofe, 2013; 김성우·정건섭, 2010; 김소연·김영호, 2013).

$$y = \beta X + u, \quad u = \lambda Wu + \epsilon \quad [\text{식 2-8}]$$

공간적 효과 중 이질성을 반영하기 위한 방법으로 제안된 GWR은 지역별로 변수 간의 관계가 다르다는 점에 착안하여 각 지역별 회귀식을 산출

하는 국지적 모델이다(Fotheringham et al., 2002). 이는 한 관측 지점, $i(u_i, v_i)$ 에서부터 일정한 범위(공간가중행렬, W) 내에 발생한 사건만을 대상으로 회귀식을 도출(식 2-9)한다는 점에서 각 지점별로 다른 회귀계수를 도출할 수 있다. 회귀계수를 도출하는 과정(식 2-10)에서 공간가중행렬이 추가되어, 공간계량모델과 같이 공간적 의존성이 반영된다는 점 역시 이 모델의 장점이라고 볼 수 있다. 지리가중회귀를 활용한 헤도닉 모델은 주택 가격에 영향을 주는 요인들이 지역에 따라 그 영향력이 다를 수 있어 널리 활용되고 있다(Fotheringham et al., 2002; Bárcena et al., 2014; 강창덕, 2010; 진찬우·이건학, 2014). 이는 지역에 따라 주택이 갖는 특성에 대한 선호의 정도가 다르다는 것을 입증하는 것이다. 특히 진찬우·이건학(2014)의 연구에서는 지역에 따라 주택 가격을 구성하는 요소들의 집합 자체가 다를 정도로 공간적 이질성이 크다는 것을 보이고 있다.

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \epsilon_i \quad [\text{식 2-9}]$$

$$\hat{\beta}_i = (X^T W_i X)^{-1} X^T W_i y \quad [\text{식 2-10}]$$

최근 헤도닉 모델은 공간뿐만 아니라 시간적인 영향도 고려하는 방향으로 발전하고 있다. 공간적 자기상관성은 물론 시간적 자기상관을 동시에 다룰 수 있는 시공간자기회귀모델(spatiotemporal autoregressive model: STAR)(Cressie, 1993; Cressie and Wikle, 2011)을 활용한 헤도닉 모델이 제시되었다(Pace et al., 1998; Sun et al., 2005; 박헌수, 2003; 전해정·박헌수, 2014). 이 모델은 공간계량모델에서 활용된 공간가중행렬이 시공간의 3차원으로 구성됨으로써 주변지역 뿐만 아니라 과거로부터 영향을 받는 주택 가격을 반영하고자 하였다. [식 2-11]에서 s 와 t 를 각각 공간과 시간적 근린을 설정하는 행렬, ρ_s 와 ρ_t 를 각각에 대한 자기회귀모수라고 했을 때, 시공간가중행렬 W 는 공간과 시간의 필터링의 선형 결합으로 표현되며 필터링 순서에 따라 다른 결과를 도출할 수 있으므로 시간과 공간의 두 가지 상호작용 효과를 모두 더해준다(박헌수, 2003).

$$W = \rho_S S + \rho_T T + \rho_{ST} ST + \rho_{TS} TS \quad [\text{식 2-11}]$$

또한 국지적 모델에서 역시 시간적 효과를 반영하고자 하는 시도가 있었다. Huang *et al.*(2010)은 시공간가중회귀모델(geographically and temporally weighted regression: GTWR)을 제시하고 이를 주택 시장에 적용함으로써 보다 설명력 높은 모델을 도출할 수 있었으며 공간적 이질성이 시간에 따라 변함을 보였다. 이는 국지적인 주택 시장에서 선호하는 주거 서비스가 변할 수 있음을 의미할 수 있으나, 단위가 다른 시간과 공간 축을 결합하는 방법상의 문제를 해결하는데 초점을 맞춘 연구(Fotheringham *et al.*, 2015b)로서 이에 대한 추가적인 연구가 필요하다.

이상을 요약하면, 미시적 측면에서 개별 주택 가격을 추정하는 헤도닉 모델은 주택이 갖는 공간적 효과를 반영하고자 공간계량모델이나 지리가중회귀 등을 통해 공간 헤도닉 모델로 발전하였다. 최근 시간적으로 누적된 데이터를 과거에 비해 쉽게 가용할 수 있어짐에 따라 시간적 효과를 반영한 시공간 헤도닉 모델도 제기되어 이를 통해 시간의 흐름에 따른 변화를 반영하여 미래 주택 가격을 예측하고자 하였다,

그러나 횡단면적인 접근을 시도하는 헤도닉 모델에서 동태적 변화를 포착하기 쉽지 않다. 특히 주택의 전세와 매매가격 간 시공간적 상호작용을 포착하고자 하는 연구에서 시공간 헤도닉 모형은 적합하지 못하다. 1절에서 살펴보았듯이 전세와 매매가격은 장기적 균형 관계를 유지하고자 변하지만 헤도닉 모형은 기본적으로 공급이 고정되어있는 한 시점에서의 균형만을 설정하기 때문이다. 이에 따라 시간적 효과가 모델에 포함되더라도 이는 시간에 의한 자기상관 효과가 필터링 된 것일 뿐 변수 간 동태적 상관관계를 파악하기에는 한계가 있다.

2) 시계열 모델의 활용 및 한계

앞서 살펴본 헤도닉 모델이 미시적 관점에서 주택 가격을 횡단면적으로 추정하는 방법이라면, 거시적 관점에서는 주택 시장 가격의 시계열적 변화를 야기하는 변수 혹은 가격들 간의 상관관계를 종단면적으로 알아보려고 한다. 이 때 주로 시계열 분석(time series analysis) 방법이 활용된다. 시계열 분석은 매우 고전적인 방법으로서 과거의 값이 현재에 영향을 미치는 시간적 자기상관관계(autocorrelation)에 주목하였다. 가장 간단한 형태의 자기회귀모델(autoregressive model; AR)을 살펴보면 다음의 [식 2-12]와 같다.

$$Y_t = \phi Y_{t-1} + u_t \quad \text{[식 2-12]}$$

이는 시차가 1인 AR(1) 모델으로, Y_t 는 현재의 관측값, Y_{t-1} 는 과거의 관측값, u_t 는 백색잡음(white noise)로 평균이 0이고 분산이 일정한 확률적 오차를 의미한다. 현재의 관측값 Y_t 가 다른 외생변수의 영향 없이 오직 해당 변수의 과거값에 의해 결정된다고 보는 것이다. 만약 현재의 관측값이 확률적 오차의 누적된 합으로 구성된다면 이는 이동평균모델(moving average: MA)으로 볼 수 있다(식 2-13).

$$Y_t = u_t - \theta_1 u_{t-1} \quad \text{[식 2-13]}$$

Box and Jenkins(1976)은 자기회귀모델과 이동평균모델을 결합하여 보다 ARIMA(autoregressive integrated moving average)모델을 제시하였다. 이는 시계열적으로 정상성(stationarity)을 띄지 않는 자료에 대한 적분과정을 포함하여 $ARIMA(p,d,q)$ 로 표현하며 다음의 [식 2-14]와 같이 나타낸다. 이 때, 시계열적 정상성이란 평균과 분산, 공분산이 시간에 따라 변하지 않는 상태를 의미한다. 그러므로 $ARIMA(p,d,q)$ 모델에서 d 는 d 차분 후 시계열이 정상적으로 변하는 값을 의미하며, p 와 q 는 각각 자기

회귀모델과 이동평균모델의 시차이다.

$$Y_t = \mu + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + u_t - \theta_1 u_{t-1} - \dots - \theta_q u_{t-q} \quad [\text{식 2-14}]$$

이상의 모델들은 단변량 시계열 분석 방법으로 이를 다변량으로 확장한 벡터자기회귀모델(vector autoregression: VAR)이 제시되었다(Sims, 1980). 이는 다른 변수의 과거값 역시 현재의 특정 변수에 영향을 미칠 수 있다는 아이디어를 담고 있다. 시차가 1인 이변량 VAR의 예시를 살펴보면 다음의 [식 2-15]와 같다. 변수 y_1 은 y_1 과 y_2 의 과거값 모두에 의해 결정되는 식으로 표현되며 이를 [식 2-16]와 같이 벡터 형태로 나타낼 수 있어 이를 벡터자기회귀모델이라고 칭한다. 이를 k 개의 변수를 갖고 p 차의 시차를 갖는 일반식으로 확장한 VAR(p)모델은 다음의 [식 2-17]로 표현된다.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \quad [\text{식 2-15}]$$

$$Y_t = \mu + AY_{t-1} + \epsilon_t \quad [\text{식 2-16}]$$

$$Y_t = \mu + \sum_{i=1}^p A_i Y_{t-i} + \epsilon_t \quad [\text{식 2-17}]$$

VAR 모델은 단순한 추정과 변수의 내생성에 대한 임의적인 구분이 필요 없으며 연립방정식 모델에 비해 정확한 예측 결과를 도출한다는 점에서 가장 성공적인 모델로 각광받으며(Hsu et al., 2008), 여러 분야에서 많이 적용되었다(Kinal and Ratner, 1986). 그러나 한편으로 추정계수의 직접적인 해석이 어려워 추가적인 분석을 통해 변수 간 상호작용 패턴을 파악하고자 하였다. 대표적으로 충격반응분석과 분산분해가 존재한다.

충격반응분석(impulse response analysis)이란 내생변수로 구성된 하

나의 구조(system)에서 외부적인 효과가 한 내생변수에 작용하였을 때, 전체 구조가 어떻게 반응하는지를 살펴보는 것이다. 단순한 이변량 VAR 모델(식 2-15)에 적용하여 생각해 보면, 오차항 ϵ_{1t} 의 변화는 직접적으로 변수 $y_{1,t}$ 의 변화를 초래함과 동시에 구조 내에서 연결되어 있는 변수 $y_{2,t}$ 에도 영향을 미친다. 왜냐하면 y_2 의 현재값은 y_2 의 과거값 뿐만 아니라 y_1 의 과거값에도 영향을 받기 때문이다(문권순, 1997). 이를 일반화하여 t 시점에 가해진 j 번째 변수의 충격에 대한 s 기때 i 번째 변수의 반응량을 $\phi_{ij,s}$ 로 나타내면 아래 [식 2-18]과 같다(박헌수·김태경, 2008).

$$\phi_{ij,s} = \sum_{k=1}^s \phi_{s-k} A_k \quad [\text{식 2-18}]$$

한편 분산분해는 충격반응분석과 반대 방향에서 오차를 분석하는 방법이라고 볼 수 있다. 충격반응분석이 외부적인 충격에 의한 효과가 확산되는 과정을 보는 방법이라면, 분산분해는 자신의 총 변화를 다른 변수의 영향이 얼마만큼 설명할 수 있는지를 살펴보는 것이다. 이는 j 변수에서 h 기까지 발생한 총 예측오차 분산 중 k 변수가 차지하는 비율을 통해 확인 할 수 있으며, j 지역에서 발생한 총 오차(mean square error: MSE)는 [식 2-19]로 나타낼 수 있다. 이 때 각각의 수식 기호는 <표 2-2>를 통해 정리하였다.

$$MSE[y_{j,t}(h)] = \sum_{i=0}^{h-1} \sum_{k=1}^K (e_j^T \Theta_i e_k)^2 = \left(\sum_{i=0}^{h-1} \Theta_i \Theta_i^T \right)_{jj} = \left(\sum_{i=0}^{h-1} \Phi_i \Sigma_u \Phi_i^T \right)_{jj} \quad [\text{식 2-19}]$$

이는 각각의 변수에서 특정 시기(h 기)까지 발생한 예측오차를 구한 식으로, 이 중에서 k 변수로 설명되는 비율을 다음의 [식 2-20]과 같이 구할 수가 있다. 이 때 h 기까지 j 변수에 영향을 준 k 변수의 설명비율은 $w_{jk,h}$ 로 표현하며, 백분율로 해석할 수 있다.

<표 2-2> VAR모델의 분산분해 수식 기호(notations)

기호	의미
e_j	I_{NK} 의 j 번째 열
jj	행렬 X 의 jj 번째 요소(대각요소)
Σ_u	오차항 u_t 의 공분산 행렬
P	Σ_u 의 Cholesky decomposition로 구해진 하삼각행렬, $\Sigma_u = PP^T$
Φ_i	계수행렬 B 의 i 제곱 행렬, $\Phi_i = JB^i J^T$
J	$J = [I_{NK} \ 0 \dots 0]_{NK \times NKp}$ 인 행렬
Θ_i	$\Theta_i = \Phi_i P$

$$w_{jk,h} = \sum_{i=0}^{h-1} (e_j^T \Theta_i e_k)^2 / MSE[y_{j,t}(h)] \quad [\text{식 2-20}]$$

이러한 보완적인 기법을 통해 다양한 해석이 가능해진 VAR 모델은 국내 주택 시장에서의 매매와 전세가 사이의 상호작용을 분석하는 기본 모델로 활용되었다. 왜냐하면 앞서 1절에서도 살펴보았듯이 매매가격과 전세가격 사이의 이론적인 인과관계가 명확하지 않으므로(임재만, 2013), 인과성을 가정하는 일반적인 회귀식으로 접근하기에 한계를 보이기 때문이다.

대표적으로 황두현(1990)은 한국주택은행에서 제공된 주택매매가격지수와 전세가격지수를 대상으로 1986년 1월부터 1990년 4월까지 54개월을 VAR(1) 모델을 통해 분석하였다. 이 때, 주택시장을 서울과 대도시(직할시), 중소도시로 구분하고 주택 유형별(아파트, 단독주택, 연립주택)로도 구분하여 각 하위 시장 내에서의 매매-전세가격 간 관계를 밝히고자 하였다. 임재만(2004)은 1986년 1월부터 2004년 10월까지의 서울과 강남, 강북의 매매가격지수와 전세가격지수를 시차가 2인 VAR(2) 모델에 적용하여 분석하였으며, 조주현·임정호(2004) 역시 같은 기간의 자료를 활용하였으나 월세시장을 추가하여 변수가 3개인 다변량 VAR(2) 모델을 활용하였

다. 임규채·기석도(2006)의 연구에서는 주택 유형별 하위시장으로 구분된 매매가와 전세가를 변수로 하는 복잡한 모델을 적용함으로써 아파트 매매가가 연립주택 전세가에 원인이 되는 등의 주택 하위시장에서 발생하는 매매-전세가의 상호작용을 포착할 수 있었다. 전해정(2012) 역시 2002년 9월부터 2012년 2월까지 서울 지역의 아파트 매매가격지수와 전세가격지수를 대상으로 VAR 모델을 이용하여 글로벌 금융위기 이후의 주택 시장의 변화를 이해하고자 하였다.

그러나 이러한 연구들은 공간상에서 발생하는 경제활동을 다루면서 공간적인 특성을 반영하지 못한 한계를 내포하고 있다. 즉, 주택 가격이 갖는 공간적인 상호작용 효과를 실증적으로 분석하지 못하였다. 주택 가격의 공간적 전이효과는 해외뿐만 아니라 국내에서도 이러한 시계열 분석 방법을 활용하여 입증되었다. 이에 따라 이러한 공간적 맥락을 고려한 매매-전세 가격의 상호작용에 관한 연구가 진행될 필요가 있다. 대표적으로 이용만·이상한(2004)은 VAR모형을 활용하여 강남지역의 주택 가격이 서울 강북지역과 성남, 전국 주택가격에 영향을 미치고 있음을 밝혔으며, 김경민(2007)은 강남구 주택 가격이 전국 137개 시군구에 영향을 주는지 파악하기 위해 각 시군구의 주택 가격이 Y 가 되고 강남구 주택가격이 X 가 되는 자기회귀분배시차모형(Autoregressive distributed lag model: ARDL) 137개를 분석하였다. 서승환(2007) 역시 강남의 주택 가격이 갖는 지역적 연관성에 초점을 맞추어 VAR의 정상성 가정을 완화한 벡터오차수정모형(vector error correction model: VECM)을 활용하여 분석하였다.

시계열 분석을 활용한 이상의 연구들에서는 주택 가격이 갖는 공간적인 효과를 반영하기 어려운 방법론적인 한계를 지녔다. 특히 앞서 살펴본 공간적 의존성을 다룰 수 있는 방법이 부재하였다. 이에 따라 지역 간에 발생하는 상호작용이 아닌 하위 시장 내부에서의 관계를 밝히는데 주목하거나(황두현, 1990; 임재만, 2004), 지역별 하위시장에 대한 고려가 미흡하였다(임규채·기석도, 2006). 또한 주택 가격이 갖는 공간적 파급효과를 살펴본 연구에서도 단순히 지역을 하나의 변수로만 취급하여 분석함으로써 파급효과가 과대로 추정되거나 부정확한 결과를 초래할 수 있는 한계를

내포하고 있으며 더욱이 매매가격과 임대가격 간의 공간적 상호작용을 살펴본 연구는 매우 드물다.

이에 따라 Beenstock and Felsenstien(2007)은 시계열적 분석을 공간적으로 확장한 공간적 벡터자기회귀모델(spatial vector autoregression: SpVAR)을 제안하며, 시계열 분석이 갖는 공간적 한계를 극복하고자 하였다. [식 2-21]은 공간적 벡터자기회귀모델의 가장 간단한 형태로 시차항(Y_{t-1})이 1개이고, 공간차항(Y_t^*)이 1개인 $SpVAR(1,1)$ 을 의미한다.

$$Y_t = \mu + A^* Y_t + B^* Y_{t-1} + \Theta^* Y_t^* + \Lambda Y_{t-1}^* + \epsilon_t \quad [\text{식 2-21}]$$

이 모델은 t 시기의 특정 지역 n 에서의 k 변수 값은 현재의 다른 변수와 과거의 다른 변수들 그리고 현재 주변지역으로부터 가중된 변수와 과거의 가중된 값으로 설명될 수 있음을 표현하는 것으로, 변수 간의 관계를 시공간적으로 정의한다. 그러므로 SpVAR은 특정한 현상의 시공간적 상호작용 효과를 알아보는데 적합하다. 모델을 제시한 Beenstock and Felsenstien(2007)은 SpVAR 모델을 활용하여 1987년부터 2004년까지 예루살렘 인근 9개 지역의 주택 가격과 소득, 인구, 주택 재고량의 관계를 살펴본 결과, 예루살렘의 소득 상승이 주변지역 주택 상승에 영향을 미침을 보였다. 이는 공간차항이 추가되어 이전 방법들이 간과한 공간적 자기상관 효과를 명백히 반영할 수 있어 공간적 파급 효과가 과대 추정되지 않는다는 장점을 보였다. 또한 이러한 방법을 차용한 Kuethe and Pede(2011)은 미국 서부지역 11개 주들 간에 발생하는 주택 가격의 순환적 구조를 밝혔다.

그러나 SpVAR 모델은 복잡해진 구조는 물론 공간 단위 별 속성을 하나의 변수로 취급한다는 점에서 짧은 시계열의 길이(관측값)에 비해 많은 변수를 갖는 문제가 있다. 즉, 자유도의 문제가 발생함에 따라 추정되는 모수의 분산이 커져 추정 결과를 신뢰하기 어렵다. 기존에는 이러한 문제를 해결하고자 변수별로 통합하여 하나의 단위(block)을 형성하고 이들 간의 관계를 규명하였다. 이에 따라 모든 지역에서의 변수 간 관계는 전역적

인 형태와 같다는 등질성 가정을 하게 된다. 그러나 앞서 1절에서 살펴본
앞듯이 주택 가격은 지역에 따라 변수 간의 관계가 다르게 형성되는 이질
성을 보임에 따라 이러한 가정은 비현실적이며, 부정확한 추정 결과를 도
출할 수 있다는 한계를 지닌다.

이에 따라 본 연구에서는 SpVAR 모델이 갖는 자유도 문제를 원자료
그대로 활용하는 추정방법을 제안함으로써 해결하고자 한다. 변수의 개수
가 관측치의 수보다 많을 경우 자료의 통합 없이 문제를 해결할 수 있다
면, 비현실적인 가정을 추가하지 않아도 되며 개별 공간 단위에서 해석이
가능할 것이다. 그러므로 본 연구에서는 축소추정법의 일종인 Lasso를
SpVAR 모델의 추정 방법으로 제안하고자 한다. 이에 따라 다음 장에서
SpVAR 모델을 도출하는 과정과 추정방법 상의 문제를 확인한 후 이를
해소할 방법을 제안하고자 한다.

제 3 장 SpVAR-Lasso 모델링

본 장에서는 앞서 언급한 SpVAR 모델을 보다 구체적으로 살펴보고자 한다. 이를 위해 우선적으로 시공간 모델링 기법인 SpVAR을 적용하기 위한 자료들의 시공간적인 특성에 대한 탐색적 방법들을 살펴볼 것이다. 시간과 공간에 의한 자기상관효과를 필터링 하는 SpVAR은 기본적으로 분석하고자 하는 자료의 시공간적 효과가 발생할 경우 적합한 방법이다. 모델의 공간차향과 시차향을 통해 자료에서 발생할 수 있는 오차 문제를 해결해주므로, 만약 자료에서 시공간적 효과가 없을 경우 보다 단순한 모델을 적용할 수 있기 때문이다. 즉, 시공간적 효과에 의해 자료가 기존의 단순한 모델들로 설명이 되지 않을 경우 SpVAR 모델의 적용이 필요한 것이다. 이에 따라 1절에서는 주택 가격에서 발생할 수 있는 시공간적 효과를 시계열적 정상성과 공간적 의존성 및 공간적 이질성의 개념을 중심으로 측정하고 평가하는 방법을 제시하고자 한다.

2절에서는 기존의 SpVAR의 추정 방법의 작동 원리와 한계를 파악하고, Lasso를 접목하여 개선된 SpVAR 모델(SpVAR-Lasso)을 제시하고자 한다. 기존의 SpVAR 모델은 관측값의 개수에 비해 큰 변수의 개수 문제를 해결하고자 자료를 변수별로 묶은 뒤 집합 간의 추정을 실시하였다. 이 과정에서 변수들의 관계가 모든 지역에서 동질적(homogeneous)이라는 가정을 전제하여 전역적 모델을 도출하였다. 그러나 특히 주택 가격 간의 관계는 지역에 따라 다르게 나타나는 이질성을 띄고 있어, 국지적으로 다른 패턴을 포착하기에 한계가 있었다. 전역적 모델과 달리 Lasso를 활용한 추정은 변수의 압축을 요구하지 않으므로, 원자료를 그대로 활용한 추정이 가능하다. 이에 따라 지역별로 다르게 형성되는 변수 간의 관계를 조망할 수 있는 국지적인 모델링이 가능하다. 이후 3절에서는 개선된 모델인 국지적 모델의 추정과 예측의 정확성을 평가하고 오차의 독립성을 검정하는 방법을 제시함으로써 SpVAR-Lasso의 방법론적인 장점을 도출하고자 한다.

제 1 절 주택 가격의 시공간적 특성 탐색

본 절에서는 시공간 모델인 SpVAR을 적용하기 위한 자료의 사전적 탐색 기법들을 확인하고자 한다. 즉, 분석하고자 하는 대상의 시공간적 특성을 확인함으로써 시공간적 모델을 활용할 수 있는지 여부를 확인하는 과정이다. 만약 분석 자료가 시간적 또는 공간적 효과가 존재하지 않다면, 보다 단순한 OLS, SAR, VAR 모델을 적용하는 것이 효율적일 수 있기 때문이다. 이에 따라 시간적인 특성은 시계열적 정상성 개념을 중심으로 살펴보고, 이를 확인할 수 있는 방법으로 수정정보완된 딕키-풀러(Augmented Dickey-Fuller, ADF) 검정과 패널 단위근 검정을 제시한다. 마지막으로 비정상성(non-stationary)을 처리할 수 있는 방법을 확인한다. 또한 공간적인 특성은 앞서 2장에서 언급한 두 가지 공간적 효과를 중심으로 살펴볼 것이다. 첫 번째는 공간적 의존성으로서 전역적 수준에서 공간적 자기상관을 측정하는 방법들을 제시하고, 이 중 Moran's I가 갖는 장점을 활용하고자 한다. 두 번째 공간적 이질성은 국지적 Moran's I와 GWR를 통해 국지적 수준에서 발생하는 공간적 자기상관의 차이와 공간적으로 다르게 형성되는 변수 간 관계를 확인하고자 한다.

1) 주택 가격의 시계열적 정상성

시계열적 정상성이란 시간의 흐름과 무관하게 평균과 분산이 일정한 것으로 시계열 분석에 있어 오래된 이슈 중에 하나이다. 왜냐하면 분산이 일정하지 않을 경우 시계열 분석을 통해 구한 예측 방정식이 부정확해지는 오류가 발생하기 때문이다. 또한 시계열 분석이 많이 활용되는 거시경제 자료들은 일정하게 증가하는 추세를 갖는 경우가 빈번하기 때문에 시계열적인 정상성을 확인하고 이를 처리하는 연구 방법이 다양하게 발전하였다(Beenstock and Felsenstein, 2010).

우선 개별 시계열의 정상성을 판단하는 방법으로 가장 널리 알려진 것은 ADF 검정(Dickey and Fuller, 1987)이다. 이는 해당 시계열의 단위근

이 존재하는지 여부를 확인하는 것으로 다음의 [식 3-1]로 표현되는 단일 시계열 자료가 불안정적 확률과정인지를 확인하는 것이다.

$$Y_t = \rho Y_{t-1} + u_t \quad [\text{식 3-1}]$$

[식 3-1]에서 Y_t 가 특정변수의 t 시점의 관측값, Y_{t-1} 은 $t-1$ 기의 관측값, u_t 는 시간의 변화에 무관한 백색잡음(white noise)라고 한다. 이 때 계수값 ρ 가 1일 경우, 해당 시계열은 시간에 따라 분산이 증가하게 되어 시간이 흐름에 따라 분산이 일정하다는 안정적 확률과정(stationary stochastic process)가정에 위배되게 된다. 그러므로 초기의 딕키-플러 검정(DF)은 ρ 가 1이 아님을 증명한다. 구체적으로 이를 추정하고자 양변에 Y_{t-1} 을 빼줌으로 1차분하여, 새롭게 구해진 계수 δ 가 0이라는 귀무가설($H_0: \delta = 0$)을 기각하는지 여부를 판단하였다(식 3-2).

$$\begin{aligned} Y_t - Y_{t-1} &= \rho Y_{t-1} - Y_{t-1} + u_t \\ \Delta Y_t &= \delta Y_{t-1} + u_t, (\text{단}, \delta = \rho - 1) \end{aligned} \quad [\text{식 3-2}]$$

수정보완된 딕키-플러 검정은 오차항 u_t 의 상관관계가 있을 경우 발생하는 추정상의 편의(bias)를 보완하기 위해 차분의 시차변수를 추가하였다. 이를 상수항과 추세항이 포함된 일반적인 형태로 표현할 경우 다음의 [식 3-3]과 같이 표현되며, [식 3-2]와 마찬가지로 δ 의 값이 0이 아님을 통계적으로 유의한 수준에서 검증한다.

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^m \alpha_i \Delta Y_{t-1} + \epsilon_t \quad [\text{식 3-3}]$$

그러나 이러한 검증 방법은 단일한 시계열, 즉 변수가 1개인 자료에 대해 적합한 방법이다. 여러 시계열이 결합된 자료인 패널 데이터에서 각각의 시계열에 대한 ADF 검정을 시행할 경우, 분산이 다양한 시계열 자료

로 인해 통계 검정력이 약해지기 때문에 패널 데이터에서는 적합하지 않다(신현중, 2006). 이에 따라 패널 데이터의 단위근 여부를 확인하고자 다양한 시도들이 진행되었으나(Quah, 1994; Levin and Lin, 1993), 변화가 등질적인(dynamic homogeneous) 패널의 단위근 검정만 가능하였다. Im *et al.*(2003)가 제안한 새로운 패널 단위근 검정 방법은 패널을 구성하고 있는 개별 시계열의 ADF t-통계량의 평균을 근거로 검정을 시행한다. 이때, 귀무가설은 모든 지역 i 의 시계열이 단위근을 갖고 있음으로 설정된다(식 3-4). 귀무가설이 기각되게 되면 일부 지역에서 정상적인 시계열 분포를 보이고 있음을 증명한다.

$$\begin{aligned}\Delta Y_{i,t} &= \alpha + \beta_i Y_{i,t-1} + \epsilon_{i,t} \\ H_0 : \beta_i &= 0, \forall i\end{aligned}\quad [\text{식 3-4}]$$

검정을 위한 통계량(식 3-5)은 각 시계열로부터 도출된 t-통계량(t_{i,T_i})으로부터 평균과 분산을 구하며, 이를 몬테카를로 시뮬레이션으로 구한 모평균(μ)과 모분산(σ^2)과의 비교를 통해 통계적으로 유의한지 검정한다. Im *et al.*(2003)의 시뮬레이션 결과, 지역이 50개이고 시계열 길이가 50인 패널 자료(상수항과 추세항이 있는 경우)의 유의수준 5%에서 임계값(critical value)은 -2.36으로 나타난다.

$$t-bar = \frac{1}{N} \sum_{i=1}^N t_{i,T_i} \quad [\text{식 3-5}]$$

이상의 방법들을 통해 다변량 시계열 자료의 비정상성이 발견될 경우, 모든 시계열에 비정상성이 사라질 때까지 동일하게 차분을 시행하거나 비정상성을 그대로 활용하는 모델을 활용한다. 차분 방법은 매우 간단하지만, 차분 횟수가 증가할수록 변수의 정보량(사례 수)이 감소하며 해석이 어렵다. 반면 오차수정모델(vector error correction model; VECM) 등 비정상적 시계열을 직접 활용하는 경우, 자료의 손실 없이 분석할 수 있다

는 장점이 있으나, 각 시계열 변수 간의 공적분의 측정과 임계치에 대하여 논쟁이 남아 있다(Gujarati and Porter, 2012).

2) 주택 가격의 공간적 의존성 및 이질성

앞선 2장 2절에서 살펴본 공간적 효과를 상기하면, 공간적 의존성 (spatial dependency)이란 공간 데이터가 갖고 있는 고유한 특성 중의 하나로(Anselin, 1988), 주변지역의 값이 특정지역의 값에 영향을 미치는 현상으로 정의할 수 있다. 어떠한 대상의 입지의 상대성으로 인해 발생하는 효과로서 관측 단위의 임의성, 공간적 합역의 문제, 공간적 외부성과 이전 효과 현상의 공간적 상호작용 등에 의해 나타난다. 이러한 관점에서 주택 가격은 공간적 의존성을 가질 가능성이 높다. 왜냐하면 한 지역에 동시에 개발되는 경우가 많아 가격에 영향을 미치는 근린 요소를 공유하면서도 주변 지역에 미치는 외부효과를 발생시키기 때문이다. 또한 결정적으로 실질적인 주택 가격을 책정하는 부동산 업체들이 책정 가격의 기준을 주변 시세로 정하기 때문에 주택 가격은 공간적으로 의존적일 수밖에 없다 (Fotheringham *et al.*, 2015a).

이러한 공간적 의존성은 공간적 자기상관(spatial autocorrelation)으로 측정이 가능하다. 대표적으로 Moran(1950)의 I지수와 Geary(1954)의 C지수가 있다. Moran의 I지수는 특정 지역 i 의 관측값 y_i 를 평균 \bar{y} 와 비교하여, 주변 지역의 값 y_j 가 평균에 비해 어떠한지를 볼 수 있는 통계량이다 (식 3-6).

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{(\sum_{i=1}^n \sum_{j=1}^n w_{ij}) \sum_{i=1}^n (y_i - \bar{y})^2} \quad [\text{식 3-6}]$$

Moran의 통계량 I는 -1에서 +1사이의 값을 갖는데, 특정 지역에 평균

이상(또는 이하)에 값이 밀집한 경우 1에 가까운 수를 가지며, 주변 지역은 평균 이하이지만 특정 지역의 값은 평균 이상일 경우 -1에 가까운 값을 갖는다. 이 때, +1에 가까운 경우를 양의 자기상관(positive autocorrelation)이라고 하며, 비슷한 값들이 군집해있는 경우를 뜻한다. 반대로 -1에 가까운 음의 자기상관(negative autocorrelation)은 서로 다른 값들이 군집한 경우이며 절대값이 0에 가까울수록 임의적인(random) 공간 분포를 갖고 있음을 의미한다. 이는 일반적인 상관관계를 나타내는 Pearson의 r과 유사한 해석이 가능하여 직관적인 이해가 가능하다는 점에서 널리 활용되고 있다.

한편 Geary의 C지수는 Moran 지수가 특정 값의 크기를 반영하지 못하는 한계를 극복하기 위해 개발되었다. Moran 지수는 특정한 값이 공간적으로 밀집해있는지 여부만을 보여줄 뿐, 모여 있는 값이 높은 값인지 낮은 값인지는 알 수 없기 때문이다. Geary의 C지수는 0에서 2사이의 값을 가지며, 평균과의 비교가 아닌 개별 값들 간의 비교를 통해 구하므로 높은 값이 밀집할수록 2에 가깝게 나타난다(식 3-7).

$$C = \frac{(n-1) \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - y_j)^2}{2 \left(\sum_{i=1}^n (y_i - \bar{y})^2 \right) \left(\sum_{i=1}^n \sum_{j=1}^n w_{ij} \right)} \quad [\text{식 3-7}]$$

이러한 공간적 자기상관 지수들은 일변량에 대한 지표로 한 변수의 공간적인 분포가 스스로 주변지역에 영향을 주고 있는지를 파악하는 것이다. 그러나 실제 공간 상에서는 주변지역의 다른 변수의 값이 영향을 미치고 있는 경우가 빈번하다. 예를 들어 밀집한 상관관계를 보이는 주택 가격과 실업률은 공간상에서 이웃 지역의 실업률이 해당 지역의 주택 가격과 상관될 수 있다(Kueth and Pede, 2011). 이러한 관계를 측정하기 위한 지수로는 Lee(2001)의 L지수가 있다(식 3-8). 이는 변수 간 상관관계를 측정하는 Pearson's r의 국지적 지표로 이해할 수 있으며(김영호, 2011), 연구 지역에서 두 변수의 상관관계가 높게 나타나는 지역이 밀집해 있는

지 여부를 확인할 수 있다.

$$L_{xy} = \frac{n}{\sum_i \left(\sum_j w_{ij} \right)^2} \frac{\sum_{i=1}^n \left[\left(\sum_{j=1}^n w_{ij} (x_j - \bar{x}) \right) \times \left(\sum_{j=1}^n w_{ij} (y_j - \bar{y}) \right) \right]}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad [\text{식 3-8}]$$

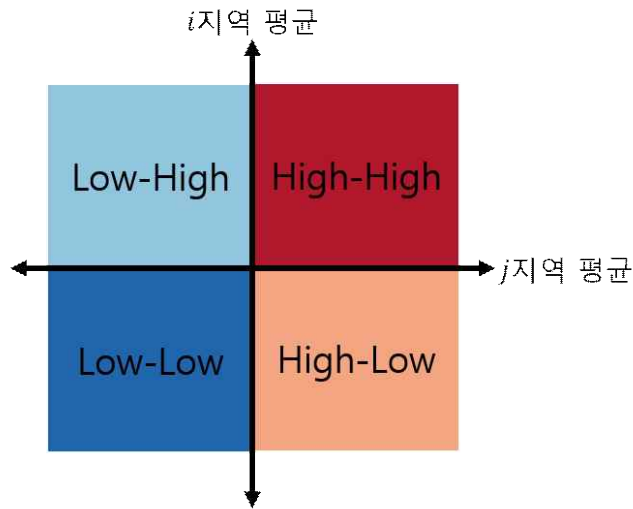
이상의 공간적 자기상관지표들은 몬테카를로 시뮬레이션을 통해 통계적인 유의성 검정이 가능하다. 공간상에 임의적으로 값들을 분포시키는 시뮬레이션을 반복적으로 수행하여 얻어진 평균값과 분산값을 바탕으로 생성된 정규분포상의 확률을 근거로 귀무가설인 공간적으로 자기상관이 발생하지 않음을 기각할 수 있는지를 평가한다.

한편 공간적 이질성(spatial heterogeneity)은 공간상에서 변수들의 관계가 지역에 따라 다르게 나타나는 현상을 의미하며, 통계적으로는 지역에 따라 변수의 평균과 분산이 다르게 나타나는 공간적 비정상성(spatial non-stationary)을 뜻한다(Bailey and Gatrell, 1995). 공간 모델링에서 이러한 이질성이 무시될 경우 추정된 결과는 추정 모수에 편의(bias)가 발생하거나, 유의수준이 잘못 설정되기도 하며 준최적(suboptimal)의 예측을 하게 된다. 주택 가격의 경우 지역에 따라 선호하는 가치가 다를 수 있고, 비탄력적인 수요와 공급으로 인해 발생하는 주택 하위시장의 존재 여부(Goodman and Thibodeau, 2003)로 인해 변수들이 공간적에 따라 이질적으로 구성될 수 있다.

이러한 공간적 이질성은 두 가지 방법으로 확인이 가능하다. 첫 번째 방법은 전역적인 통계량을 각 지역별로 나누어 살펴보는 것이다. 앞서 살펴본 공간적 의존성 역시 국지적인 수준에서 개별적인 측정이 가능하다(Anselin, 1995). Moran의 지수를 국지적으로 접근한 LISA(local indicator of spatial association)을 살펴보면 다음의 [식 3-9]와 같다.

$$I_i = \frac{n^2(y_i - \bar{y}) \sum_{j=1}^n w_{ij}(y_j - \bar{y})}{(\sum_{i=1}^n \sum_{j=1}^n w_{ij}) \sum_{i=1}^n (y_i - \bar{y})^2} \quad [\text{식 3-9}]$$

특정 지역 i 에서의 공간적 자기상관을 측정하는 이 지수는 [식 3-6]과 달리 지표에 하첨자 i 가 있어 특정 지역 i 가 고정되어 있을 때, 주변지역의 관측값과 평균의 차이를 측정한다. 이 지수를 활용하게 될 경우 지역을 네 가지로 구분할 수 있는데, 평균을 기준으로 높은 값 주변에 높은 값이 밀집해있는지(High-High), 대상지역은 평균보다 높으나 주변지역은 낮은지(High-Low)인지를 파악할 수 있다(그림 3-1).



[그림 3-1] LISA에 의한 지역 구분 모식도

더 나아가 변수들이 공간상 이질적인 평균과 분산을 갖게 됨에 따라 변수 간의 관계가 지역에 따라 다르기도 한다. 이러한 관계를 전역적인 방법으로 추정하게 될 경우 전술하였듯이 지역에 따라 모델에 의한 결과를 신뢰할 수 없을 수도 있다. 이렇듯 공간적인 이질성은 실제 공간상에서 이뤄지는 활동들에 대해 매우 일반적이다. 예를 들어 위계가 있는 현상의 경

우, 모든 지역에서 변수 간 관계가 같다고 보기 어려우며, 시간에 따라 현상이 확산되거나 소멸되는 경우도 발생하기 때문이다(Anselin, 1988). 이에 따라 국지적인 수준에서의 각기 다른 변수간의 관계를 추정하는 모델이 고안되었는데, 대표적으로 지리가중회귀(Geographically weighted Regression: GWR)가 있다.

지리가중회귀(Brunsdon *et al*, 1998)는 관측 지역마다 각기 다른 회귀식을 산출하는 방법으로 관측 지역을 중심으로 일정한 범위 내의 관측값만을 대상으로 회귀식을 산출한다(식 3-10).

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \epsilon_i \quad [\text{식 3-10}]$$

이 때 (u_i, v_i) 는 관측 지점 i 의 좌표값을 의미하며, $\beta_0(u_i, v_i)$ 는 상수항을, $\beta_k(u_i, v_i)$ 는 i 지점에서의 변수 x_k 가 y 에 미치는 영향력을 나타내는 회귀계수를 ϵ_i 는 그 지역에서 해당 식으로 설명하지 못하는 오차항을 뜻한다. 각 지역에서의 회귀계수는 일반적인 추정법인 OLS와 유사하지만 근린지역에 대한 공간가중치가 포함된다는 점에서 차이를 보인다(식 3-11).

$$\hat{\beta}_i = (X^T W_i X)^{-1} X^T W_i Y \quad [\text{식 3-11}]$$

공간가중행렬 W 는 공간적 의존성 개념에서 출발된 것으로 가까이 위치하는 것이 더 큰 영향을 미친다는 전제 하에 여러 가지 방법으로 설정된다. 대표적으로 일정거리를 설정하여 그 이내에 속한 지역만을 특정지역의 회귀식에 포함시키는 방법과 거리의 제곱의 역수만큼 가중치를 부여하는 방법이 있다. 이 때 근린의 범위는 중요한 이슈로서 선형적인 근거에 의해 고정된 범위를 근린으로 설정하는 방법(fixed kernel)도 존재하지만, 국지적인 모델의 특성상 전역적 모델에 비해 사례수가 적어져 특정 지역은 과소표본으로 인한 큰 분산이 발생하는 문제가 야기된다. 이에 따라 지리가중행렬은 근린의 폭(bandwidth)를 조율해가며 일정 수 이상의 사례를 포

함시키면서 분산을 줄이는 범위를 가중치로 설정하는데, 추가되는 관측치에 따라 변화하는 AIC(Akaike information criterion)값을 기준으로 근린의 범위를 조정(adaptive kernel)할 수 있다(Fotheringham *et al.*, 2002).

이상을 통해 SpVAR 모델을 적용하기 위한 자료의 시공간적 특성을 확인하는 방법을 알아보았다. 시공간적 효과를 필터링 해주는 SpVAR 모델을 적용하기 위한 전제 조건으로 분석 자료는 시간적으로 정상성을 띄어야 하며, 공간적으로 의존성을 띄어야 한다. 공간적 이질성은 기존의 SpVAR 모델에서 간과되었지만 본 연구가 제안하는 SpVAR-Lasso에서는 매우 중요한 요소로서 지역에 따라 이질적인 국지적 모델을 구축하는 근거가 된다.

제 2 절 주택 매매-전세가격의 시공간 상호작용 모델

본 절에서는 SpVAR 모델의 추정 방법을 살펴보고 기존 추정 방법의 한계를 Lasso를 통해 극복하고자 한다. 또한 추정 방법 간의 비교 방법을 제안함으로써 보다 정확한 모델을 채택하는 방안을 제시하고자 한다. 시계열 분석에 공간적 속성이 추가된 SpVAR은 매우 복잡하고 다차원적으로 구성됨에 따라 과소표본 문제가 나타나기 쉽다. 실세계에서 집계되는 자료의 특성상 시계열의 길이가 긴 자료(오랜 기간 수집된 자료)는 매우 드문 반면, 이를 설명하는 변수 또는 같은 현상을 집계하는 지역은 많기 때문이다. 이러한 문제들에 관하여 기존의 SpVAR은 변수별로 합쳐진 블록을 SUR(seemingly unrelated regression) 기법을 활용하여 각각의 계수를 추정하였다(Beenstock and Felsenstein, 2007; Kuethe and Pede, 2011). 그러나 이 과정에서 지역 간 자료가 변수 별로 묶임에 따라 공간적 등질성을 가정하게 된다. 공간적으로 이질적이지 않은 자료의 경우 이와 같은 방법이 보다 효과적일 수 있으나, 많은 공간 자료는 이질성을 무시하기 어렵다. 그러므로 본 연구에서는 과소표본 문제를 효과적으로 해결

하면서 동시에 매우 많은 변수들 중 유의한 변수를 효과적으로 채택하는 Lasso를 활용하여 SpVAR의 계수를 추정하는 방법을 제안하고자 한다. 이는 변수를 축소하지 않아도 되므로 비현실적인 가정을 피할 수 있어 보다 현실에 가까운 추정과 예측이 가능하다는 장점이 있다. 이를 보다 실증적으로 검토하기 위하여 모델에 활용된 기간의 자료들을 바탕으로 실시한 추정의 정확성 평가와 분석기간 이후의 데이터에 대한 예측의 정확성을 평가하며, 이때 발생된 오차의 공간적 자기상관을 측정하고 비교함으로써 모델 추정 결과의 신뢰성을 확보하고자 한다.

1) SpVAR-SUR: 전역적 모델

2장에서 소개된 SpVAR 모델을 다시 소개하면 다음의 [식 3-12]와 같다. 이 때, 각각의 기호는 <표 3-1>을 통해 정리하였다.

$$Y_{knt} = \mu_{kn} + \sum_{i=1}^K (\alpha_{ki} Y_{int} + \beta_{kj} Y_{int-1} + \theta_{ki} Y_{int}^* + \lambda_{ki} Y_{int-1}^*) + \epsilon_{knt} \quad [\text{식 3-12}]$$

<표 3-1> SpVAR의 수식 기호(notations)

기호	의미
k	변수 $k = 1, 2, \dots, K$
n	지역 $n = 1, 2, \dots, N$
t	시기 $t = 1, 2, \dots, T$
Y	관측값
Y^*	공간 가중된 관측값, $Y^* = WY$
μ	상수항
α	일반 회귀 계수
β	시계열 회귀 계수
θ	공간모델 회귀 계수
λ	지연된 공간회귀 계수(lagged spatial lag coefficient)
ϵ	오차항

결국 SpVAR은 t 시기의 특정 지역 n 에서의 k 변수 값은 현재의 다른 변수와 과거의 다른 변수들 그리고 현재 주변지역으로부터 가중된 변수와 과거의 가중된 값으로 설명될 수 있음을 표현하였다. 이를 행렬형태로 전환하여 보다 간단하게 표기하면 [식 3-13]과 같이 나타낼 수 있다.

$$Y_t = \mu + A^* Y_t + B^* Y_{t-1} + \Theta^* Y_t^* + \Lambda Y_{t-1}^* + \epsilon_t \quad [\text{식 3-13}]$$

이 때, Y_t 는 지역의 수인 n 으로 구분된 $NK \times 1$ 인 관측값 벡터이며, 각 변수의 계수는 차원이 $NK \times NK$ 인 블록 대각 행렬(block diagonal matrix)이다. 그러나 [식 3-13]에서는 우변의 Y_t 와 Y_t^* 가 오차항과 독립적이지 않게 되어 이대로 추정할 경우 회귀 계수의 추정치가 불편추정량이 아니게 되므로 그 값을 신뢰하기 어렵다. 또한 이는 추정해야 하는 모수의 수($4K^2$)에 비해 식의 개수(NK)가 부족하여 과소식별(under identification)문제가 발생하므로 축약형으로 전환하여 추정해야 한다(식 3-14).

$$Y_t = \Gamma_0 + \Gamma_1 Y_{t-1} + \Gamma_2 Y_{t-1}^* + u_t \quad [\text{식 3-14}]$$

$$\begin{aligned} M &= (I_{NK} - A^*)^{-1} \\ \Gamma_0 &= (I_{NK} - M\Theta^* W^*)^{-1} M\mu, \Gamma_1 = (I_{NK} - M\Theta^* W^*)^{-1} MB^*, \\ \Gamma_2 &= (I_{NK} - M\Theta^* W^*)^{-1} M\Lambda^*, u_t = (I_{NK} - M\Theta^* W^*)^{-1} M\epsilon_t \end{aligned}$$

이러한 축약형 방정식을 통해 도출된 회귀계수들은 직접적인 해석이 불가능하다. [식 3-14]에서도 확인할 수 있듯이 각각의 계수에는 다른 차원의 변수들의 계수가 포함되어 있는 형태이므로 도출된 숫자 그대로 이해하기는 어렵다. 다만, 축약형 계수가 통계적으로 유의하다는 사실은 상당히 큰 정보를 제공한다. SpVAR의 경우, 통계적으로 $\Gamma_1 = 0, \Gamma_2 \neq 0$ 인 경우 해당 변수의 시계열적인 변화가 직접적인 과거값보다는 공간적으로 연계된 과거값에 의한 영향이 존재하는 것을 의미함으로써, 해당 시스템이

시공간적 역동성(spatio-temporal dynamics)을 갖고 있음을 통계적으로 검증할 수 있기 때문이다(Beenstock and Felenstein, 2007).

이렇게 축약된 방정식의 경우 각각의 변수에 대해 독립적인 회귀식으로 추정이 가능하나 시계열의 길이가 충분하지 못한 경우, 이를 하나의 블록으로 묶어 추정한다. 이 때, 각 블록에 속한 관측치의 개수는 지역별로 통합되므로 NT 개로 과소표본 문제가 해결되어, 낮은 자유도에 의해 발생할 수 있는 높은 분산에 의한 예측 부정확성을 해소한다. 그러나 이 과정에서 공간상에서 이질적으로 나타나는 변수 간의 관계를 포착하지 못하게 되므로 등질성을 가정하고 추정을 실시한다. 다만, 축약된 SpVAR 모델에서 각 블록 안의 식들은 각 지역별 관계식들이므로 지역 간 발생하는 공간적 상관관계(spatial correlation)가 오차항에서 발현될 수 있기 때문에, 통상 최소자승법 대신 SUR(seemingly unrelated regression)을 사용한다는 것이 특징적이다.

Zellner(1962)에 의해 제안된 SUR은 자유도가 제한된 다수의 회귀식에서 다른 식들의 오차항 간에 상관성이 있는 경우 적절한 추정 방법이다(Anselin, 1988). 만약 여러 개의 회귀식에서 설명변수가 모든 식에서 동일하다면, SUR의 추정 결과는 OLS와 같은 결과를 도출하지만, 그렇지 않은 경우에는 일반최소자승법(general least square; GLS)으로 추정된 결과가 불편추정치가 된다. SpVAR의 각 변수별 블록은 지역별 회귀식으로 구성되어 있으며, 각 회귀식은 각 지역별 설명변수로 이뤄져있다. [식 3-15]는 임의의 이변량 SpVAR모델의 한 블록을 제시한 것으로 각각의 회귀식은 해당 지역의 변수만을 설명변수로 포함한다. 이에 따라 다른 지역의 변수에 의한 영향(예시에서 $X_{2,t-1} \rightarrow Y_{1,t}$ 관계를 의미)은 오차항을 통해서만 설명이 가능하다는 한계를 지니고 있다.

$$\begin{aligned}
 Y_{1,t} &= \mu_1 + \alpha_1 Y_{1,t-1} + \beta_1 Y_{1,t-1}^* + \gamma_1 X_{1,t-1} + \delta_1 X_{1,t-1}^* + \epsilon_1 \\
 Y_{2,t} &= \mu_2 + \alpha_2 Y_{2,t-1} + \beta_2 Y_{1,t-1}^* + \gamma_2 X_{2,t-1} + \delta_2 X_{2,t-1}^* + \epsilon_2 \\
 &\dots \\
 Y_{N,t} &= \mu_N + \alpha_N Y_{N,t-1} + \beta_N Y_{N,t-1}^* + \gamma_N X_{N,t-1} + \delta_N X_{N,t-1}^* + \epsilon_N
 \end{aligned} \quad [\text{식 3-15}]$$

회귀계수에 의한 설명이 부족하다는 문제를 해소하고자 제안된 SpVAR의 충격반응함수는 다음의 식 전개를 통해 이해할 수 있다(Beenstock and Felsenstein, 2007). 우선 [식 3-12]에 변수가 1개인 경우를 적용한 행렬식은 [식 3-16]과 같다(상수항 제외).

$$Y_t = \beta I_N Y_{t-1} + \theta W Y_t + \lambda W Y_{t-1} + \epsilon_t \quad [\text{식 3-16}]$$

이를 후행연산자 L 을 활용하여 정리하면 [식 3-17]가 도출된다.

$$\begin{aligned} (A + BL)Y_t &= \epsilon_t \\ A &= I_N - \theta W \\ B &= -(\beta I_N + \lambda W) \end{aligned} \quad [\text{식 3-17}]$$

[식 3-17]을 Y_t 에 대한 식으로 정리하기 위해 양변을 같은 항으로 나누어 주면, 오차항의 변화에 따른 시스템을 이루고 있는 내생변수 Y_t 의 변화를 나타내는 식으로 표현된다(식 3-18).

$$\begin{aligned} Y_t &= C^{-1} \epsilon_t + \sum_{i=1}^N a_i r_i^t \\ C &= (A - BL) \end{aligned} \quad [\text{식 3-18}]$$

여기서 a_i 는 초기 조건에 의한 상수항을 의미하며, r_i 는 각 지역(N)별 변수(K)의 고유값(eigenvalues)들로 [식 3-19]를 통해 구할 수 있다.

$$|C^{-1} - r I_N| = 0 \quad [\text{식 3-19}]$$

만약 [식 3-19]에서 모든 r_i 의 절대값이 1보다 작다면 $\sum_{i=1}^N a_i r_i^t$ 이 0으로 수

럼하게 되어, 특정 지역에 외부로부터 가해진 충격은 C^{-1} 만큼 다른 지역으로 확산 될 것이다. 이 때, $|r_i| < 1$ 은 주어진 자료가 시공간적으로 정상적(stationary)임을 의미한다. 이를 활용하여 다른 지역에 가해진 외부적 충격(초기 조건)에 의해 변하는 특정 지역에서의 내생변수 $Y_{i,t}$ 는 다음의 [식 3-20]으로 구할 수 있다.

$$Y_{i,t} = \frac{\epsilon_{i,t} - \pi\epsilon_{i,t-1} + \sum_{j=1, i \neq j}^N (\theta\epsilon_{j,t} + \lambda)\epsilon_{j,t-1}}{(1-r_iL) \prod_{j=1, i \neq j}^N (1-r_jL)} + \sum_{i=1}^N A_i r_i^t \quad [\text{식 3-20}]$$

결국 SpVAR 모델에서 다른 지역의 변화는 θ 만큼의 관계를 갖고 있는 주변 지역에 누적적으로 영향을 미치며, λ 만큼 주변 지역의 과거값에 연쇄적으로 영향을 미치게 된다.

2) SpVAR-Lasso: 국지적 모델

앞선 항에서는 SpVAR의 기본적인 추정 방법을 살펴보았다. 기존의 추정 방법은 사례수가 적을 경우(시계열의 길이가 짧을 경우) 발생할 수 있는 자유도의 문제로 인해 변수별로 자료를 통합함으로써 공간적인 등질성을 가정하게 된다. 이에 따라 다른 지역의 다른 변수에 의한 영향을 오차항에 간접적으로 파악할 수 있을 뿐 모델 내부에서 직접적으로 알기 어렵다. 그러나 현실 공간에서 발생하는 대부분 공간 자료는 공간적 의존성뿐만 아니라 이질성을 함께 갖고 있다. 또한 시계열적으로 오랜 기간 집계된 자료가 많지 않아, 관측값보다 변수의 개수가 많을 경우가 많다. 따라서 보다 현실적인 추정을 위해서는 기존의 추정 방법이 갖고 있는 연산 상의 한계를 극복해야 할 필요가 있다. 즉, 주어진 자료를 통합하거나 변수를 선택적 또는 임의적으로 제거함으로써 발생할 수 있는 왜곡을 막기 위해 모든 자료를 활용하면서도 효율적으로 유의한 변수를 선택해야 한다. 이를

위한 방법으로 축소추정법의 하나인 Lasso를 SpVAR 모델의 추정 방법으로 제안하고자 한다.

축소추정법(shrinkage estimation)이란 회귀식에서 관측값의 개수(n)가 변수의 개수(p)보다 작은 경우($n < p$)의 비직교성(non-orthogonal) 문제 해결에 관한 방법이다(Hoerl and Kennard, 1970). 임의의 편의(bias)를 벌점(penalty)으로 부과함으로써 추정되는 회귀계수($\hat{\beta}$)을 축소시키는 이 방법은 벌점화 회귀(penalized regression)라고도 불린다. 이를 살펴보고자 우선적으로 앞서 2장에서 제시한 일반적인 회귀식과 통상최소자승법을 상기하고자 한다(식 2-5과 식 2-6).

$$y = X\beta + \epsilon \quad [\text{식 2-5}]$$

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad [\text{식 2-6}]$$

이 때, X 는 p 개의 변수로 구성된 변수 행렬($n \times p$)이고 p 개의 변수별 회귀계수 β 는 $p \times 1$ 인 벡터이다. 통상최소자승법은 추정값과 관측값의 차이로 발생하는 잔차의 제곱합을 최소화하는 방법인데, 잔차가 정규성, 등분산성, 독립성 가정(independent and identically distributed; i.i.d.)을 만족할 경우의 회귀계수는 최량불편추정치(best most efficient linear unbiased estimator; BLUE)가 된다.

그러나 이와 같은 추정은 변수의 개수가 관측치의 값보다 많은 경우 또는 변수들 간의 상관관계가 높아 다중공선성(multicollinearity)이 발생하는 경우 적용이 불가능해지거나 추정된 결과를 신뢰하기 어렵다. 왜냐하면 [식 2-6]에서의 $X^T X$ 의 역행렬이 존재하지 않는 문제(특이행렬 문제)가 발생하기 때문이다. 즉, 특이행렬(singular) 문제가 발생함에 따라 하나의 해가 도출되지 않으므로 회귀계수를 추정할 수 없게 된다. 축소추정법이 개발되기 이전에는 이러한 문제를 해결하기 위해 임의적으로 변수를 제거하거나 요인분석 등의 차원 축소법을 통해 변수의 개수를 조정하였다. 그러나 이러한 방법은 연구자에 임의적인 선택이 개입된다는 점에서 중립성을

유지하기 어렵고 축소된 차원은 변수의 의미를 파악하기 어렵하다는 한계를 내포하고 있다(Hoerl and Kennard, 1970).

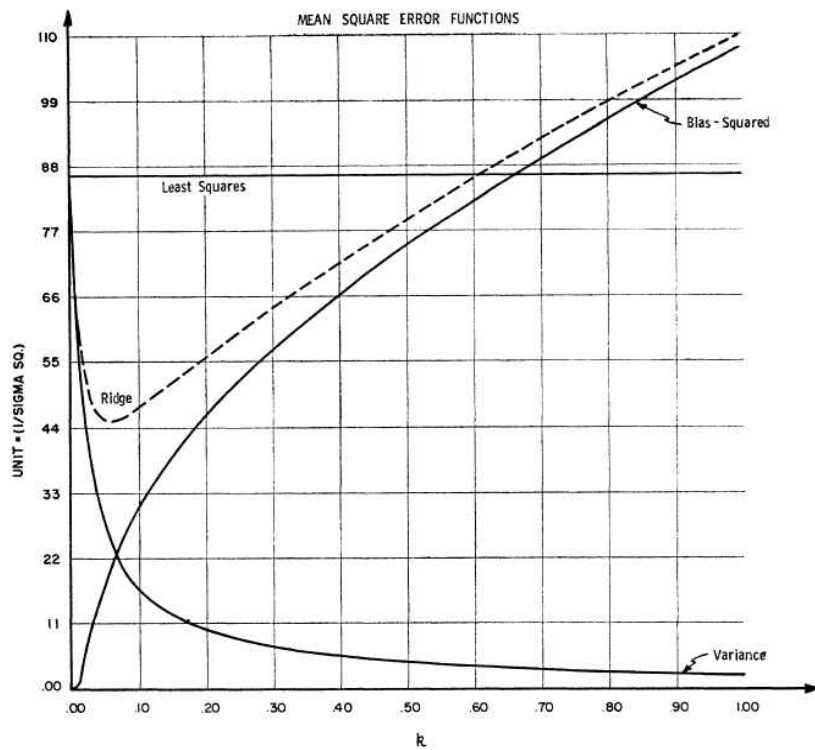
이러한 문제를 해결하기 위해 제시된 능형회귀(ridge regression)는 최소제곱법이 가정하고 있는 불편추정 가정을 완화시켜 더 나은 모델을 추구한다. 편의와 분산 사이 존재하는 상쇄관계를 활용하여 일정량의 편의를 추가함으로써 보다 오차가 작은 결과를 도출할 수 있기 때문이다(식 3-21). 역행렬이 존재하지 않는 $X^T X$ 에 총오차를 최소화할 수 있는 임의의 수 k 를 대각행렬에 더해줌으로써 가역행렬로 만드는 방법이다. 이 때 k 는 조율모수로 [식 3-22]의 제약식을 만족시키는 라그랑지안 승수이다(박창이 외, 2011).

$$\hat{\beta}^{ridge} = (X^T X + kI)^{-1} X^T Y \quad [\text{식 3-21}]$$

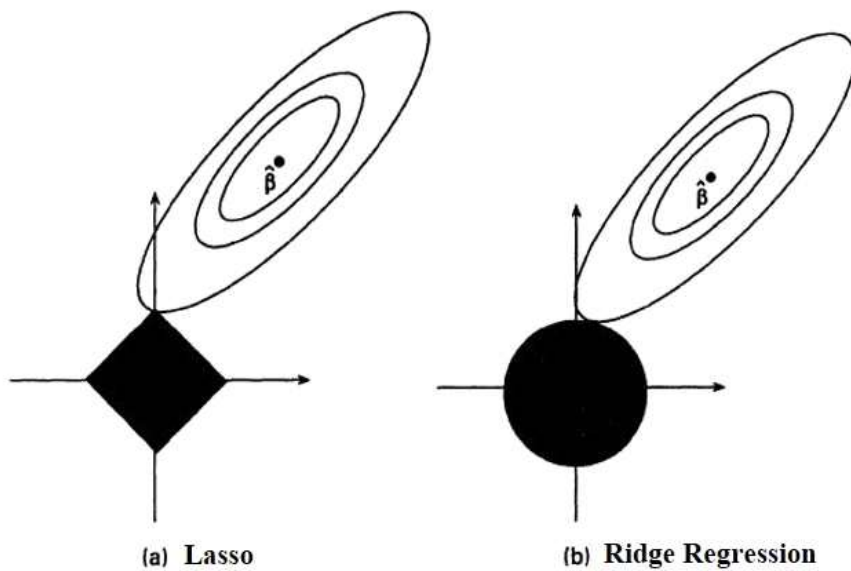
$$\begin{aligned} \hat{\beta}^{ridge} &= \operatorname{argmin}_{\beta} \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 \\ \text{subject to} & \quad \sum_{j=1}^p \beta_j^2 \leq t^2, t \geq 0 \end{aligned} \quad [\text{식 3-22}]$$

[그림 3-2]은 능형회귀식에서 발생하는 편의와 분산 간의 상쇄관계를 보여주는 그래프로 조율모수 k 가 상승함에 따라 편의는 증가하며 분산은 감소된다. 이 때 통상최소자승법은 조율모수가 없는 경우를 의미함에 따라 k 축과 평행한 평균잔차제곱(mean square error: MSE) 값을 보인다. 반면, 능형회귀는 k 값이 일정수준까지 증가할 때 오차가 감소하는 경향을 보인다. 이러한 원리로 인해 조율모수 k 가 적절하게 선택된다면, 능형회귀는 통상최소자승법보다 효율적인 모델을 제시할 수 있게 된다.

그러나 능형회귀의 경우 변수에 편의를 부과하여 보다 효과적인 추정을 가능하게 하지만 특정한 변수를 선택하는 기준은 존재하지 않는다. 이후 등장한 Lasso는 능형회귀에서는 제약조건을 바꿈으로써 모델에서 유의하지 않은 변수를 0으로 수렴시키는 방법을 포함하였다(Tibshirani, 1996).



[그림 3-2] 통상최소자승법과 능형회귀의 평균잔차제곱
출처: Hoerl and Kennard, 1970



[그림 3-3] 추정 회귀계수의 기하학적 이해
출처: Tibshirani, 1996

[식 3-22]와 달리 제약식 부분이 절대값으로 되어 있음(식 3-23)에 따라 회귀계수가 0이 될 수 있는 가능성(우도, likelihood)이 높아지게 된다. 이를 기하학적으로 설명한 [그림 3-3]를 살펴보면, 변수가 2개인 가상의 공간에서 t^2 인 능형회귀의 제약식 영역은 원형으로 나타난다. 반면 Lasso의 제약식 영역은 $|t|$ 이므로 정사각형으로 표현된다. 그림에서 동심원 모양의 타원체로 구성된 회귀계수의 우도 함수는 오차를 최소화 하면서 제약식을 만족해야 한다(식 2-20, 식 2-21). 이에 따라 회귀계수의 추정값 $\hat{\beta}$ 은 제약식과 우도함수의 접점에서 형성되는데, Lasso의 경우 그 접점이 한 변수의 절편에서 형성되는 반면 능형회귀의 경우 절편에서 형성될 확률이 극도로 낮아지게 된다. 결국 Lasso에서는 한 변수의 회귀계수가 0이 될 가능성이 높으며, 이는 해당 변수가 모델에서 유의하지 않음을 뜻할 수 있는 것이다. 이 때, [식 3-24]를 만족시키는 라그랑지안 승수 λ 는 일반적으로 교차검정(cross-validation)을 통해 오차를 최소화할 수 있는 값으로 설정된다.

$$\begin{aligned} \hat{\beta}^{lasso} &= \operatorname{argmin}_{\beta} \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j)^2 \\ &\text{subject to} \\ &\sum_{j=1}^p |\beta_j| \leq t, t \geq 0 \end{aligned} \quad [\text{식 3-23}]$$

$$\hat{\beta}^{lasso} = (X^T X + \lambda I)^{-1} X^T Y \quad [\text{식 3-24}]$$

유의하지 않은 변수를 0으로 수렴시키는 Lasso는 변수 선택에서 있어 모든 경우의 수를 고려하면서도 효과적으로 유의한 변수의 집합을 선택할 수 있다는 장점이 있다. 기존의 변수 선택 방법들은 변수의 입력 순서에 따라 결과가 다르거나 준최적 변수 집합일 수 있으며(전진선택법 혹은 후진선택법), 지나치게 많은 연산을 요구하는(최적화기준법) 문제를 가졌으나 Lasso는 이를 효과적으로 해결하였다(Hsu *et al.*, 2008).

과소표본 문제와 변수 선택 문제를 비교적 간단하게 해결할 수 있는 Lasso는 다양한 회귀 모델에 접목되어 활용되고 있다(Zhu *et al.*, 2004;

Hsu *et al.*, 2008; Wheeler, 2009; Friedman *et al.*, 2010). Zhu *et al.*(2004)는 종속변수가 0과 1로 구성된 로지스틱 회귀에서의 Lasso를 활용하는 방안을 제시하였으며, Friedman *et al.*(2010)은 일반화선형모델 (generalized linear model)에 Lasso를 적용하여 보다 일반적인 회귀식에서 Lasso 추정이 가능한 방법을 고안하였다. 한편 Wheeler(2009)와 Hsu *et al.*(2008)은 각각 공간과 시간 모델에 Lasso를 접목하였다. 전자의 경우 공간적 국지 모델인 지리가중회귀(geographically weighted regression; GWR)에서 발생할 수 있는 다중공선성 문제를 Lasso를 활용하여 해결하는 방안을 모색하였다. 후자는 시계열 방법론 중 벡터자기회귀모델에 Lasso 추정법을 대입했을 때, 기존의 VAR이 갖는 자유도 문제를 해결함과 동시에 기존의 추정 방법보다 정확하고 효율적인 결과를 도출하며 특히 시계열의 길이가 짧을 때 그 효과가 더욱 극명함을 보였다.

Hsu *et al.*(2008)의 연구에서 밝혀진 VAR 모델과 Lasso 추정의 결합 가능성을 활용하여, VAR의 공간적 확장버전인 SpVAR 모델에서도 Lasso 추정을 적용하고자 한다. SpVAR의 일반식 [식 3-13]을 단순선형모델 형태로 전환하기 위해 [식 3-25]와 같이 변수를 재설정하였다. X 는 관측값으로 구성된 벡터로 $2NK+1 \times 1$ 차원의 열벡터이며, B 는 계수값으로 구성된 $NK \times 2NK+1$ 차원의 행렬이다(식 3-26).

$$\begin{aligned} Y &= BX^* + u_t \\ Y &= ((X^*)^T \otimes I_{NK})\beta + u_t \equiv X\beta + u_t \end{aligned} \quad [\text{식 3-25}]$$

$$\begin{aligned} Y &= (Y_1, \dots, Y_t)^T, X^* = (1, Y_{t-1}, Y_{t-1}^*)^T, \\ B &= (\Gamma_0, \Gamma_1, \Gamma_2), \beta = \text{vec}(B) \end{aligned} \quad [\text{식 3-26}]$$

최소자승추정의 경우 잔차의 제곱합(residual sum of square; RSS)을 최소로 하는 β 값을 추정하지만, Lasso를 비롯한 벌점화 회귀는 잔차제곱에 제약식을 추가(식 3-27)하여 이를 만족시키는 최소의 회귀계수를 추정한다(Hoerl and Kennard, 1970). Lasso의 경우 능형회귀와 달리 제약식

부분이 절대값으로 구성되어 있어 유의하지 않은 변수를 0으로 수렴시킨다는 것을 2장에서 살펴보았다. SpVAR 모델에서도 마찬가지로 j 번째 회귀계수인 β_j 가 유의하지 않다면 0으로 수렴된다. 이 때 λ 는 조율모수(tuning parameter) s 의 라그랑지안 승수로 0일 경우 OLS와 같은 추정 결과를 도출한다(박창이 외, 2011)(식 3-28).

$$\begin{aligned} \tilde{\beta} &= \arg \min_{\beta} (Y - X\beta)^T (Y - X\beta) \\ \text{Subject to} & \\ \sum |\beta_j| &\leq s, s \geq 0 \end{aligned} \quad [\text{식 3-27}]$$

$$\begin{aligned} \tilde{\beta} &= \arg \min_{\beta} (Y - X\beta)^T (Y - X\beta) + \lambda \sum |\beta_j| \\ \tilde{\beta} &= (X^T X + \lambda I)^{-1} X^T y \end{aligned} \quad [\text{식 3-28}]$$

이를 통해 도출되는 모델은 SUR 추정과 달리 오차의 공간적 상관관계가 나타나지 않으며, 모델 내부에서 공간적 이질성이 반영된다. 즉, 각각의 회귀식이 각기 다른 변수 집합으로 구성된 국지적 모델이다. 예를 들어, SUR 추정에서는 변수별 집단(주택 종류별 거래형태) 간의 회귀식을 구성했기 때문에 강남구 아파트의 전세가와 서초구 아파트의 매매가 간의 관계는 전역적인 수준의 전세가와 매매가 관계로 가정되었다. 하지만 Lasso 추정에서는 강남구 아파트의 전세가가 독립변수일 경우, 모든 주택 하위시장 가격이 설명변수로 투입되어도 추정이 가능하며 이 중 유의한 변수들만 0이 아닌 것으로 된다. 이에 따라 SUR 추정에서 전역적인 관계로 가정하였던 서초구 아파트 매매가의 영향력이 다르게 도출될 수 있다는 점에서 SpVAR-Lasso는 국지적 모델이라고 볼 수 있다.

이러한 Lasso 추정은 Efron *et al.*(2004)에 의해 고안된 *lars* 알고리즘(least angle regression algorithm)에 의해 OLS와 비슷한 연산 속도가 가능해져 매우 효율적이다(Hsu *et al.*, 2008). 본 연구에서는 *lars* 알고리즘에 기반을 둔 R package *glmnet*을 활용하였으며, 자료를 훈련자료(training set)와 검증자료(test set)으로 나누어 교차검정을 실시한 결과가

최소예측오차에 1 표준편차만큼 더한 값을 λ 값으로 설정함으로써 과적합 문제를 해결하고자 하였다.

3) 모델의 비교 및 평가

마지막으로 각기 다른 추정 방법에 의해 도출된 SpVAR 모델의 결과를 비교하기 위한 방법을 제시한다. 이는 앞서 제시한 SpVAR-Lasso의 성능을 평가하기 위한 것으로 첫 번째 방법은 추정과 예측 결과의 정확성 측정이다. 이 때, 추정(estimate)은 모델에 투입된 자료를 모델에 의해 다시 복원하는 과정을 의미하며, 예측(prediction)은 미래에 대한 예견을 의미한다. 이들의 정확성 비교는 평균절대오차(MAE)와 평균제곱근오차(RMSE)로 가능하다. 평균절대오차는 오차의 절대값의 평균(식 3-29)으로 비교 대상의 사례수가 동일하여 관측치에 개수에 의한 영향이 없을 때, 직관적인 해석이 가능한 지표로 널리 활용된다(Hydmand and Koehler, 2006). 평균제곱근오차는 오차 제곱의 평균으로 자료의 스케일과 무관하게 사용할 수 있는 장점이 있다(식 3-30).

$$MAE = \frac{1}{N} \sum_{i=1}^n |y_i - \hat{y}_i| \quad [\text{식 3-29}]$$

$$RMSE = \sqrt{(\sum_{i=1}^n (y_i - \hat{y}_i)^2) / N} \quad [\text{식 3-30}]$$

두 번째 평가 방법으로는 오차항의 공간적 상관관계를 살펴보는 것이다. SUR 추정의 경우 모델 자체에서는 공간적인 이질성을 포착할 수 없으므로 오차에서 발생하는 공간적 상관관계를 근거로 이질적인 관계를 간접적으로 확인한다. 만약 Lasso 추정 결과 오차에서 공간적인 관계가 발견되지 않는다면, 모델 자체에서 이질성 문제가 해결된 것으로 볼 수 있다. 이를 확인하기 위해 두 추정 방법에 의한 오차들을 앞서 살펴본 Moran 지수의 크기와 유의확률을 비교할 것이다.

제 4 장 SpVAR-Lasso 모델의 적용: 수도권 주택 가격

본 장에서는 앞서 3장에서 살펴본 방법론을 실제 사례지역인 수도권 지역에 적용하여, 실제 공간 상에서 나타나는 주택 하위시장에서의 매매와 전세가격 간의 상호작용 패턴을 파악함으로써 주택 가격 연구에서 SpVAR-Lasso의 적용 가능성을 확인하고자 한다. 이를 위해 1절에서는 수도권 주택 가격의 현황과 시공간적 특성(시계열적 정상성, 공간적 의존성, 공간적 이질성)을 살펴본다. 2절에서는 SpVAR 모델을 통해 수도권 주택 하위시장 간 가격 모델을 제시하고자 한다. 이 때, 등질성을 가정한 기존의 SUR 추정법을 통해 전역적 SpVAR 모델을 구축하고 본 연구에서 제안한 SpVAR-Lasso 모델을 이용하여 국지적 모델을 도출하고 이를 바탕으로 수도권 주택가격 간 시공간적 상호작용 패턴을 해석한다. 마지막으로 전역적 모델과 국지적 모델 간의 추정 및 예측의 정확성을 비교하고 오차의 자기상관여부를 판단하여 수도권 주택 가격의 시공간적 상호작용 패턴 분석에 보다 적합한 방법론을 제시하고자 한다.

제 1 절 수도권 주택 가격의 시공간적 특성

1) 수도권 주택 가격의 시간적 특성 평가

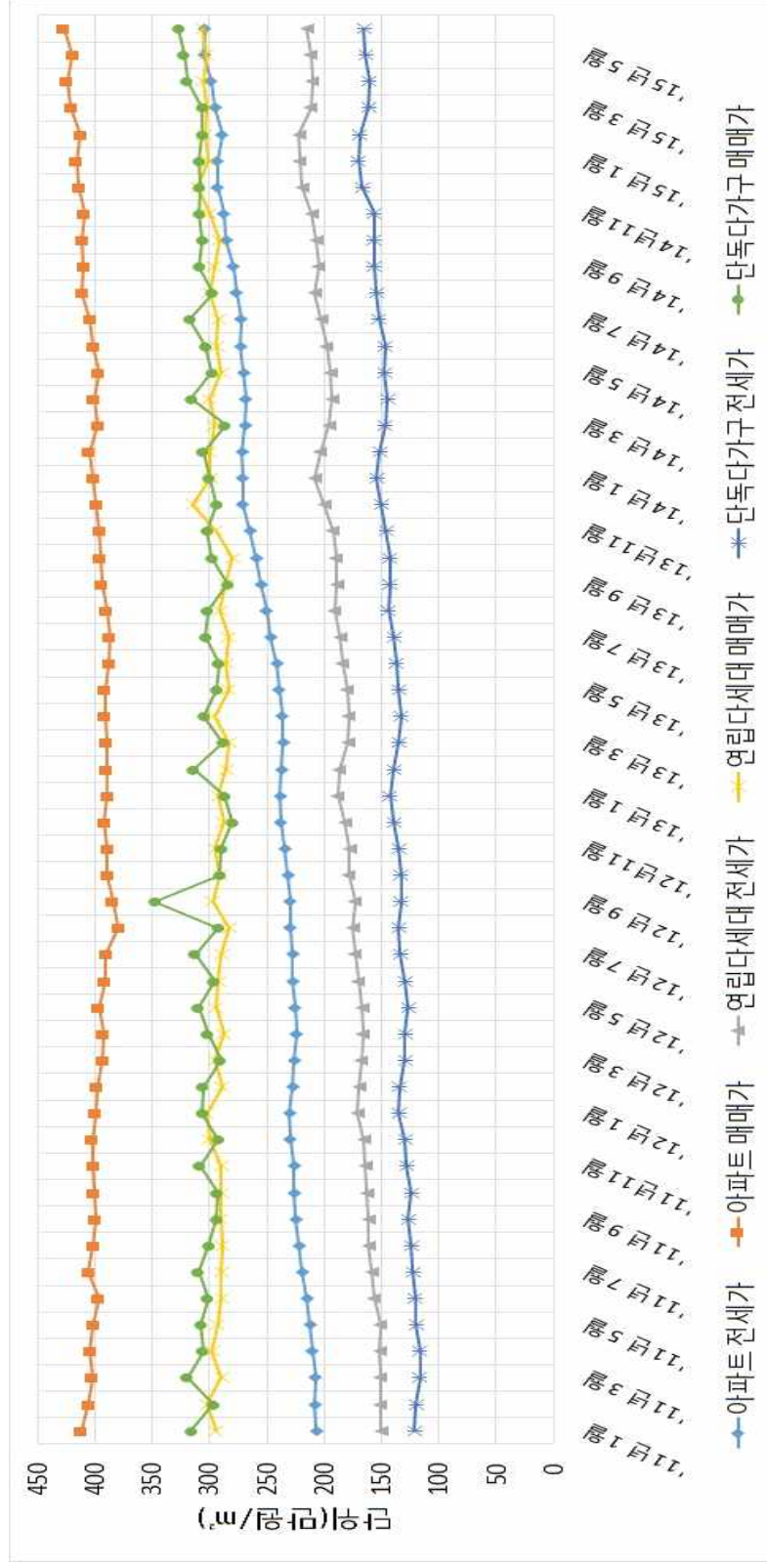
2011년 1월부터 2015년 6월까지 주택별 평균 거래가의 변화는 아래 <표 4-1>과 [그림 4-1]에서 확인할 수 있다. 2011년 이후 전체적으로 매매가보다는 전세가의 상승이 두드러졌다. <표 4-1>의 주택 종류별 전세가와 매매가의 표준편차를 살펴볼 때, 모든 주택의 경우에서 전세가의 표준편차가 보다 크게 나타났다. 이는 시계열적 평균을 기준으로 등락의 폭이 보다 크게 나타났음을 보여준다. [그림 4-1]을 통해 보다 구체적으로 살펴보면, 아파트의 매매가가 가장 높게 나타났지만 그 변동량이 미미하게 나

타난 반면, 아파트의 전세가는 다른 주택에 비해 비교적 빠르게 상승하여 연립다세대와 단독다가구 주택의 매매가와 비슷한 수준까지 올랐다. 연립다세대 주택의 경우도 비슷하게 매매가는 등락을 거듭했지만, 전세가는 꾸준히 상승하는 곡선을 보이고 있으며, 단독다가구 주택은 매매와 전세 모두 큰 변화를 보이지 않았다.

<표 4-1> 수도권 주택 종류별 거래가 평균 및 표준편차

시기	아파트		연립다세대		단독다가구	
	전세가	매매가	전세가	매매가	전세가	매매가
‘11년 6월	214.96	396.83	156.29	289.55	120.81	301.14
‘12년 6월	227.39	390.76	170.30	292.46	129.41	295.93
‘13년 6월	241.20	387.41	184.22	285.45	137.19	291.81
‘14년 6월	273.42	401.52	198.05	293.97	147.03	303.30
‘15년 6월	303.90	426.78	215.35	306.11	165.69	326.78
평균	248.88	400.13	184.15	294.18	140.42	303.25
표준편차	28.32	10.29	20.55	7.29	14.37	11.80

기간 내 주택 유형별 가격의 공간적인 분포는 일반적으로 알려진 사실과 대부분 일치하는 결과를 보인다(표 4-2~4). 흔히 강남 3구라고 불리는 강남구, 서초구, 송파구의 주택 가격이 종류와 시기에 무관하게 상위에 위치한다. 그러나 최근 서초구의 아파트 전세가가 강남구를 앞서게 되었고, 과천시의 아파트 매매가가 소위 말하는 강남3구와 유사한 수준을 유지하고 있다. 또한 연립다세대 주택의 경우 마포구, 광진구, 영등포구, 강동구 등이 강세를 보이며, 단독주택의 경우 전통적인 서울의 도심권(중구, 종로구, 용산구)에서 높은 가격을 보인다.



[그림 4-1] 수도권 주택 종류별 평균 거래가

<표 4-2> 2011년 1월 가격 기준 상위 5개 지역(만원/㎡)

순위	아파트		연립다세대		단독다가구	
	전세가	매매가	전세가	매매가	전세가	매매가
1	강남구 (438.30)	강남구 (1,093.51)	강남구 (320.87)	용산구 (776.51)	강남구 (225.89)	의왕시 (843.45)
2	서초구 (424.48)	과천시 (1,037.48)	서초구 (287.02)	서초구 (661.59)	관악구 (217.81)	성동구 (774.44)
3	송파구 (371.20)	서초구 (901.89)	송파구 (266.61)	강남구 (615.38)	서초구 (210.60)	종로구 (770.15)
4	용산구 (361.41)	송파구 (873.98)	마포구 (259.16)	과천시 (541.62)	서울중구 (204.99)	강남구 (737.49)
5	광진구 (340.72)	용산구 (817.81)	광진구 (252.89)	송파구 (519.84)	종로구 (203.56)	서초구 (727.17)

<표 4-3> 2015년 6월 가격 기준 상위 5개 지역(만원/㎡)

순위	아파트		연립다세대		단독다가구	
	전세가	매매가	전세가	매매가	전세가	매매가
1	서초구 (595.80)	강남구 (1,128.90)	강남구 (450.74)	과천시 (831.51)	강남구 (337.55)	서울중구 (1,062.38)
2	강남구 (574.32)	과천시 (1,015.73)	서초구 (432.04)	강남구 (759.62)	서초구 (302.65)	용산구 (759.72)
3	송파구 (554.41)	서초구 (1,000.29)	영등포구 (399.50)	용산구 (751.69)	종로구 (298.29)	과천시 (740.59)
4	종로구 (511.05)	송파구 (850.35)	강동구 (393.49)	영등포구 (639.35)	광진구 (256.94)	의왕시 (710.13)
5	광진구 (494.68)	용산구 (782.46)	송파구 (375.06)	서초구 (635.22)	과천시 (255.05)	강남구 (699.00)

<표 4-4> 평균 가격 기준 상위 5개 지역(만원/㎡)

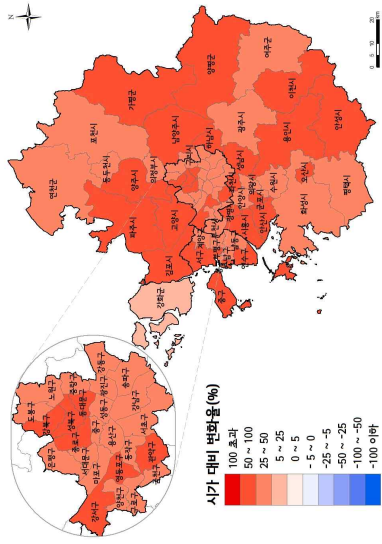
순위	아파트		연립다세대		단독다가구	
	전세가	매매가	전세가	매매가	전세가	매매가
1	강남구 (494.29)	강남구 (1,046.22)	강남구 (370.02)	용산구 (701.54)	강남구 (251.01)	서울중구 (725.30)
2	서초구 (479.85)	과천시 (904.24)	서초구 (357.50)	강남구 (637.94)	서초구 (250.38)	강남구 (681.60)
3	송파구 (450.82)	서초구 (885.70)	송파구 (325.77)	서초구 (599.99)	종로구 (230.54)	종로구 (675.51)
4	광진구 (413.61)	송파구 (841.33)	강동구 (317.53)	과천시 (590.40)	광진구 (222.89)	용산구 (660.98)
5	용산구 (407.01)	용산구 (765.71)	광진구 (312.79)	송파구 (485.17)	서울중구 (217.74)	서초구 (633.69)

한편 이와 같은 주택 가격의 변화는 공간적으로 다르게 나타난다(그림 4-2). 아파트 전세가의 경우 전 지역에서 상승을 보이고 있으나, 경부축과 서북축, 동부축을 따라 분포하는 경기 외곽지역에서의 상승이 두드러지는 반면, 매매가는 서울과 경기 북부에서 하락세를 보이는 와중에 서남축의 상승세가 돋보인다. 연립다세대주택 역시 전세가는 전 지역에서 상승을 보이고 특히 경기 남부지역의 변화가 두드러지는 반면, 매매가의 경우 경기 외곽지역에서의 변화가 크게 나타났다. 이를 요약한 <표 4-5>를 살펴보면, 서울 지역이 아닌 서울 외곽지역의 변화가 크게 나타난 것을 확인할 수 있다. 특히 인천 지역(연수구, 서구, 중구)의 상승폭이 크게 나타났으며, 경기 최외곽지역(안성시, 평택시, 이천시, 연천군)에서의 변화 역시 두드러진다.

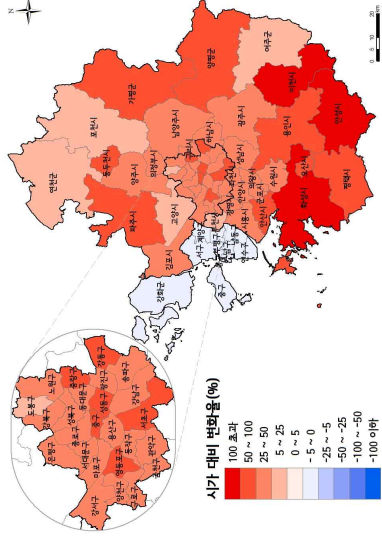
<표 4-5> 가격 변화율 상위 5개 지역(%)

순위	아파트		연립다세대		단독다가구	
	전세가	매매가	전세가	매매가	전세가	매매가
1	연수구 (85.56)	안성시 (27.63)	안성시 (119.79)	연천군 (148.26)	연천군 (143.19)	과천시 (125.9)
2	가평군 (79.21)	광명시 (26.80)	이천시 (103.73)	연수구 (88.78)	광명시 (101.53)	김포시 (91.54)
3	인천서구 (75.90)	이천시 (22.36)	화성시 (100.7)	여주시 (75.63)	구리시 (93.93)	하남시 (90.77)
4	용인시 (64.18)	평택시 (18.62)	동두천시 (84.24)	인천동구 (73.34)	인천중구 (89.19)	서울중구 (72.23)
5	종로구 (63.16)	의왕시 (18.46)	용인시 (78.78)	영등포구 (58.68)	의왕시 (87.38)	양주시 (65.27)

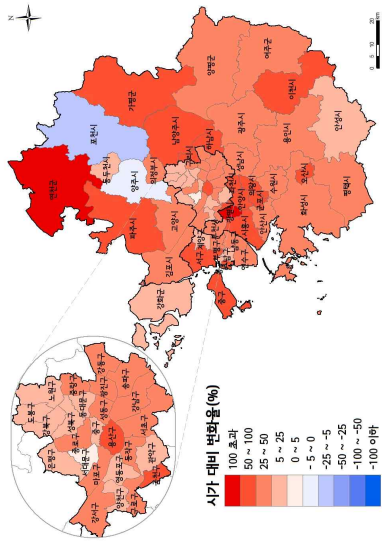
주: 변화율 = ('15년 6월 단가 - '11년 1월 단가) × 100 / '11년 1월 단가



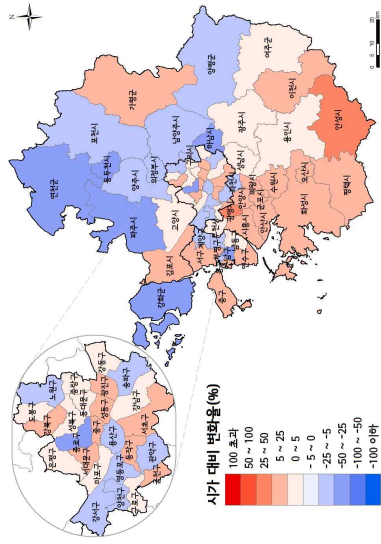
a) 아파트 전세가



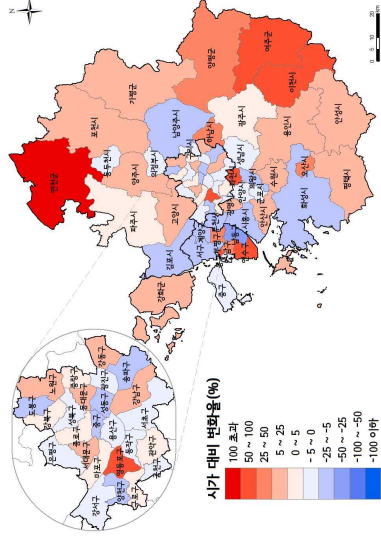
b) 연립다세대 전세가



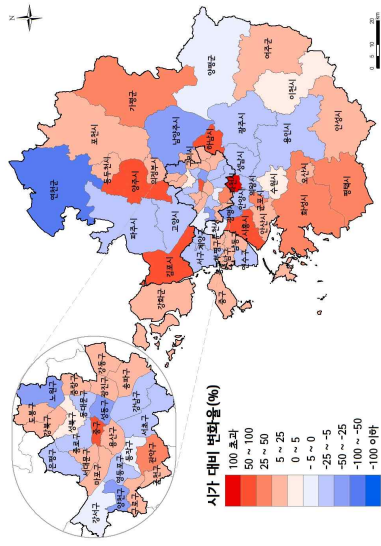
c) 단독다가구 전세가



d) 아파트 매매가



e) 연립다세대 매매가



f) 단독다가구 매매가

[그림 4-2] 수도권 주택 가격 변화율

이렇게 시간에 따라 상이한 변동을 보이는 각각의 주택 가격의 시계열적인 특성을 통계적으로 검정하기 위하여 각 개별 변수의 단위근 검정과 패널 단위근 검정을 실시하였다. 증가 추세를 보이는 시계열을 VAR 모델을 기반으로 하는 SpVAR에 적용할 경우 과대오차가 발생하며 해석상의 문제가 발생되므로 이를 사전에 검증해야 하고 자료를 변환하는 작업이 필요하기 때문이다. 이에 따라 개별 시계열의 정상성은 ADF 검정을 통해 평가하였고, 패널 단위근 검정은 Im *et al.*(2003)이 제시한 알고리즘을 기반으로 최대 시차를 8로 설정하여 실시하였다. 이 때 시차는 길수록 보다 다양한 경우의 수를 분석할 수 있지만, 8차 이후의 결과값에 큰 차이를 보이지 않아 최대 시차로 설정하였다.

우선 ADF 검정 결과, 가격 원자료의 경우 총 390개(65개 지역 × 6개 변수)의 시계열 자료 중 220개가 통계적으로 유의한 수준(10%)에서 귀무가설인 비정상성을 기각하지 못하였다. 즉, 220개의 하위시장에서의 가격 변동은 일정한 추세(주로 상승하는 추세)를 보이고 있음을 반증한 것이다. 반면 가격을 1차 차분한 가격 변화량 자료에서는 약 11개의 비정상적 시계열이 도출되었다(표 4-6). 1차 차분 자료에서의 보인 시계열적 정상성은 하위시장에서의 가격 변화율이 안정적인 것으로 해석할 수 있다. 이를 보다 구체적으로 분석하기 위해 변수별로 살펴보면, 대부분의 지역에서 아파트 가격(전세와 매매)이 비정상성을 띄는 것으로 나타났다. 특히 아파트 전세가격의 경우에는 1차 차분한 경우에도 10개 지역(경기 광명시, 경기 남양주시, 경기 의정부시, 경기 화성시, 서울 강동구, 서울 강서구, 서울 노원구, 서울 도봉구, 서울 성북구, 인천 서구)에서는 비정상성을 띄고 있다. 이는 가격 변화율이 증가하는 것으로 추후 아파트 전세가격의 변화가 극심해질 것으로 전망된다. 이례적으로 아파트 전세가격은 1차 차분한 결과도 비정상적 시계열을 보이고 있지만, 나머지 변수들의 경우 1차 차분에서 비정상성이 사라지는 결과가 나타났다.

<표 4-6> 변수별 ADF 검정 결과(비정상 시계열 개수)

변 수		가격(d = 0)	가격변화량(d = 1)
전 체		220	11
전 세	아파트	58	10
	연립다세대	25	0
	단독다가구	27	1
매 매	아파트	54	0
	연립다세대	36	0
	단독다가구	20	0

주: 유의수준 10% 기준

다음으로 실시한 패널 단위근 검정에서는 가격 원자료와 가격 변화량(1차 차분) 모두 유의한 결과를 도출하였다(표 4-7). 그러나 이를 각 변수별로 나누어 살펴보게 되면 가격 원자료의 경우 단독다가구의 전세가는 정상성 가설을 채택하지 못한다. 단독다가구의 전세가처럼 귀무가설을 기각하지 못하는 수준은 아니지만 가격 변화량 자료(1차 차분)에 비해 통계량 수치가 높은 것을 확인할 수 있다.

<표 4-7> 변수별 패널 단위근 검정 결과(t-bar statistic)

변 수		가격(d = 0)	가격변화량(d = 1)
전 체		-24.7978 ***	-126.5368 ***
전 세	아파트	13.5785 ***	-55.9022 ***
	연립다세대	1.8333 *	-45.7921 ***
	단독다가구	-0.8621	-47.8533 ***
매 매	아파트	-7.6351 ***	-57.4487 ***
	연립다세대	-25.8829 ***	-56.5670 ***
	단독다가구	-42.9608 ***	-46.7692 ***

이를 종합해보면 2011년부터 2015년 6월까지 수도권 주택 가격은 일정한 추세를 갖고 증가하는 패턴이 통계적으로 유의하게 나타났다. 전체 패널로 보았을 때 이러한 비정상성이 두드러지지는 않았지만, 아파트 전세와 같은 개별 시계열 자료에서는 추세를 무시하기 어려운 패턴을 보이고 있어 가격 자체를 SpVAR 모델에 직접 적용하기 어렵다. 다만 가격을 1차분 한 변화량의 경우 패널은 물론 개별 시계열에서 정상성을 띄고 있어 SpVAR 모델에 활용하기 적합한 것으로 판단된다. 이에 따라 이에 따라 본 연구에서는 즉 가격 변화량을 모델에 투입하고 이를 분석 대상으로하고자 한다. 즉, SpVAR 모델의 독립변수와 설명변수는 각각의 주택 하위 시장 별 가격의 변화량으로서 다음의 [식 4-1]과 같다.

$$\begin{aligned} \text{가격변화량} &= \text{당기 가격} - \text{전기 가격} \\ \Delta Y_t &= Y_t - Y_{t-1} \end{aligned} \quad [\text{식 4-1}]$$

2) 수도권 주택 가격의 공간적 효과 측정

다음으로 주택 가격 변화량의 공간적인 의존성과 이질성을 측정하여 수도권 주택 가격 변화량이 갖는 공간적 효과를 확인하였다. 공간적 의존성을 파악하기 위해 전역적인 수준에서의 변수별 Moran's I와 이변량 Moran's I를 사용하였으며, 공간적 이질성을 탐색하는 수단으로 국지적 Moran's I와 변수별 GWR 모델의 회귀계수를 살펴보았다. 분석은 각 주택별 거래의 시계열 평균 가격 변화량을 대상으로 하였으며, 공간 가중치는 경계를 접하는 지역에만 가중치를 설정하였다(식 4-2).

$$w_{ij} = \begin{cases} 1 & i \text{와 } j \text{가 경계를 접함} \\ 0 & \text{그렇지 않은 경우} \end{cases} \quad [\text{식 4-2}]$$

<표 4-8>은 각 변수별 평균 변화량의 공간적 자기상관지수로 대각행렬은 각 변수별 Moran's I를 의미하며, 하삼각행렬은 각 변수 간의 이변량

Moran's I값을 뜻한다. 이 때, 각 주택별 매매가의 변화량은 공간적 자기상관이 나타나고 있지 않지만, 전세가의 변화량은 유의한 수준의 비교적 높은 공간적인 자기상관을 나타내고 있다. 이는 전세가의 변화량이 공간상 인접한 지역에서 유사하게 나타나고 있으며, 주변지역으로 영향을 미치고 있음을 의미한다. 뿐만 아니라 아파트 전세가 상승은 주변지역의 아파트 매매와 연립다세대의 전세가, 매매가, 단독주택의 전세가 상승에 영향을 주고 있음을 볼 수 있다. 특히 아파트 전세와 연립다세대 주택의 전세는 상당한 공간적 자기상관을 보이고 있다. 예를 들어 설명하자면, 강남구 아파트 전세가의 상승은 주변 지역(서초구)의 연립다세대 주택의 전세가 상승과 함께 나타나고 있다고 볼 수 있다. 이는 두 변수 간의 시간적 선후행 관계는 알 수 없으나 공간적으로 밀접한 상호작용, 즉 공간적 동조효과를 암시하고 있다고 볼 수 있다.

<표 4-8> 변수별 공간적 자기상관지수(Moran's I)

		아파트		연립다세대		단독다가구	
		전세	매매	전세	매매	전세	매매
아파트	전세	0.5876 ***					
	매매	0.1142 **	0.2677				
연립 다세대	전세	0.5099 ***	0.1210 ***	0.5695 ***			
	매매	0.1090 **	-0.0117	0.1608 **	0.2195		
단독 다가구	전세	0.3506 ***	0.1695 **	0.3352 ***	0.089 *	0.3764 ***	
	매매	0.0404	0.0181	0.0433	-0.0112	0.0386	0.1163

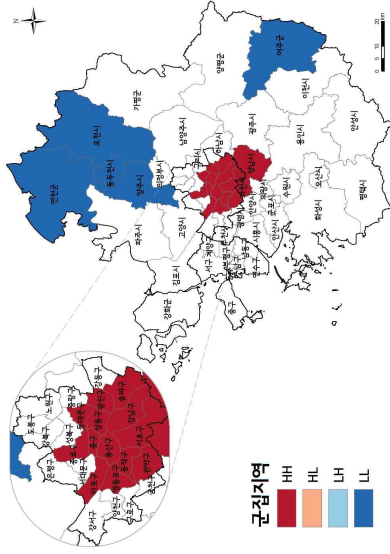
***: 99%, **: 95%, *: 90% 신뢰구간에서 유의함

마지막으로 LISA를 통해 국지적 수준의 공간적 자기상관을 확인해봄으로써 가격 변화량의 공간적인 클러스터를 파악하였다. 우선 아파트 전세가

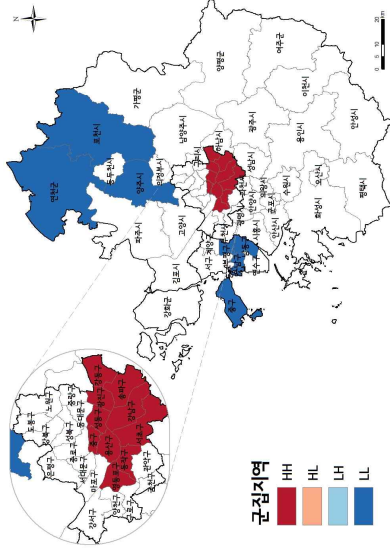
의 평균 상승폭이 높은 지역은 서울 강남구를 중심으로 밀집해 있다. 특히 서초구와 송파구, 과천시와 성남시는 기간 내 평균적으로 m^2 당 3만원 이상의 상승을 보여 수도권 평균 변화량인 $1.82\text{만원}/\text{m}^2$ 을 크게 상회하였다. 반면 경기 북부 외곽지역의 아파트 전세가는 하락하지는 않았지만 그 상승이 다른 지역에 비해 더뎠다. 양주시에서 연천군으로 이어지는 지역은 m^2 당 평균 0.5만원 미만으로 상승하여 평균에 비해 낮은 증가가 밀집한 곳으로 나타났다(그림 4-3a). 한편 전역적 수준에서의 공간적 자기상관 정도가 낮은 아파트 매매가의 경우 밀집지역이 전세가에 비해 두드러지지 않는다. 다만, 군포시와 안양시가 비교적 높은 상승이 밀집한 지역으로 나타났다으며, 서초구의 경우 주변 지역인 용산구와 과천시가 하락세를 보이는 와중에 평균($0.27\text{만원}/\text{m}^2$)보다 높게 상승하였다(그림 4-3d).

연립다세대 주택의 가격 변화량도 아파트와 비슷한 공간적인 분포를 보인다. 우선 강남권을 중심으로 높은 전세 상승이 발생한 반면 경기북부지역은 낮은 상승폭을 보인다. 그러나 아파트와 달리 인천지역 전세가의 낮은 상승이 두드러지며(그림 4-3b), 매매가 변화량의 특이한 패턴이 발견되지 않는다(그림 4-3e). 단독주택의 경우, 전세가 역시 다른 주택들에 비해 핫스팟이 용산-서초-과천으로 이어지는 지역으로 좁게 형성되었으며(그림 4-3c), 매매가의 경우 대부분 지역에서 특이한 점이 나타나지 않았다(그림 4-3f).

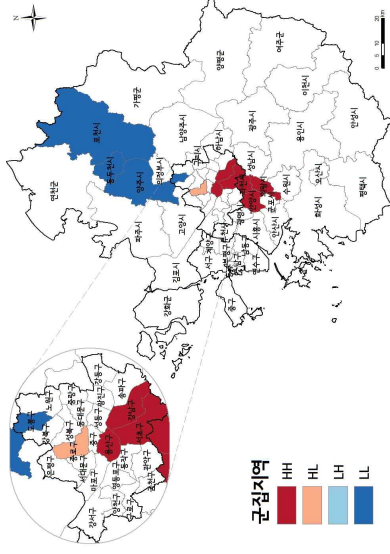
이상을 정리하면 주택별 거래가격의 변화량은 공간적으로 상호 연관관계가 있음이 통계적으로 유의한 수준에서 밝혀졌다. 전세가의 경우 모든 주택 형태에서 공간적인 자기상관이 나타나며, 특히 아파트 전세가는 다른 변수들과 긴밀한 공간적인 상관관계가 있는 것으로 나타났다. 이는 강남권을 중심으로 평균적으로 높은 상승을 보이는 아파트 전세가가 다른 지역의 변수에 영향을 미칠 수 있음을 암시하는 것으로서, 공간적 의존성을 반영하는 SpVAR 모델이 수도권 주택가격 변화량 간 상호작용 패턴 분석에 적합함을 의미한다.



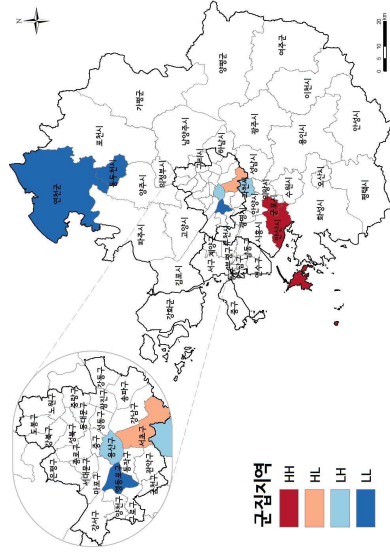
a) 아파트 전세가



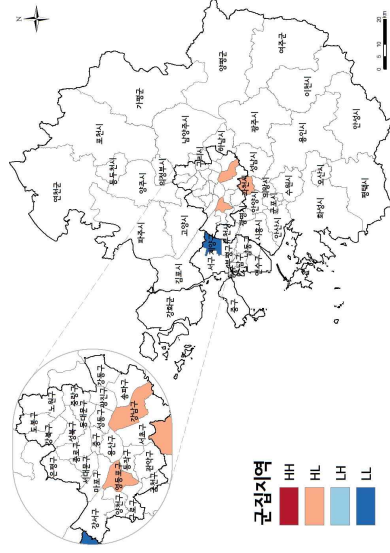
b) 연립다세대 전세가



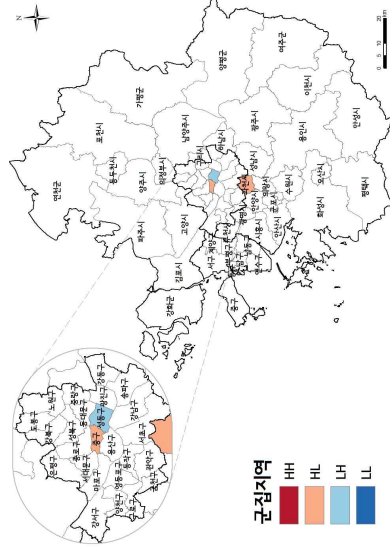
c) 단독다가구 전세가



d) 아파트 매매가



e) 연립다세대 매매가



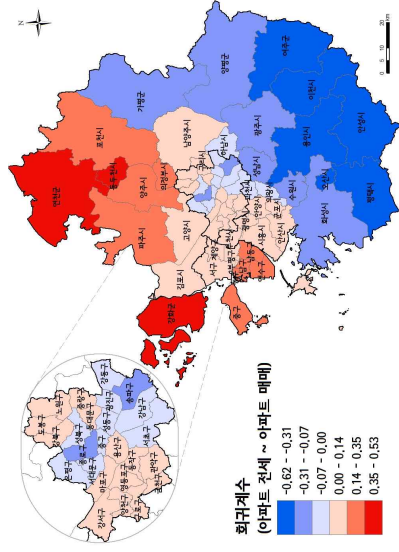
f) 단독다가구 매매가

[그림 4-3] 가격변화량의 공간적 군집

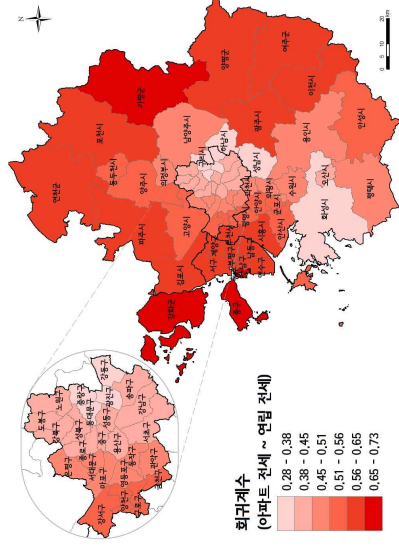
한편 공간적인 이질성을 확인하기 위한 GWR 모델은 다음의 [식 4-3]과 같이 설정하였다. 이는 i 지역의 k 번째 설명변수 Y_k 는 해당 설명변수를 제외한 모든 변수로 설명하는 것을 의미한다. 예를 들어 강남구의 아파트 전세가는 강남구의 아파트 매매가, 연립다세대의 전세가와 매매가, 단독다가구의 전세가와 매매가로 설명된다고 보았다.

$$Y_{k,i} = \alpha_{k,i} + \sum_{m=1, m \neq k}^5 \beta_{m,i} Y_{m,i} + \epsilon_{k,i} \quad [\text{식 4-3}]$$

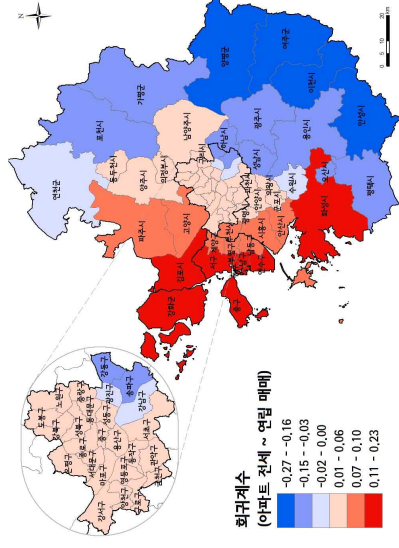
변수간의 관계가 공간적으로 다르게 구성되는 것을 [그림 4-4]를 통해 확인할 수 있다. 이는 GWR을 통해 추정된 각 변수의 회귀계수를 표현한 지도로, 같은 변수라고 하더라도 공간에 따라 다른 변수와의 관계가 정반대인 경우가 존재한다. 예를 들어 아파트 매매가와 전세가 간의 관계를 보면, 한강 이남의 지역에서는 두 변수 간 부적 관계가 형성되지만 나머지 지역에서는 정적인 관계가 존재한다. 또한 연립다세대 주택의 전세가는 전 지역에서 아파트 전세가에 긍정적인 영향을 미치지만 그 정도는 지역에 따라 상이하다. 이와 같은 결과는 수도권 주택 시장 내에서 각각의 주택별 거래가격이 지역에 따라 다르게 영향을 미칠 수 있음을 시사한다. 이에 따라 SpVAR 모델에서 가정한 공간적 동질성은 본 사례 연구에 적합하지 않을 수 있으며, 이에 따라 이질적인 관계를 살펴볼 수 있는 SpVAR-Lasso를 활용한 국지적 모델이 보다 정확한 예측 결과를 도출할 수 있음을 시사한다.



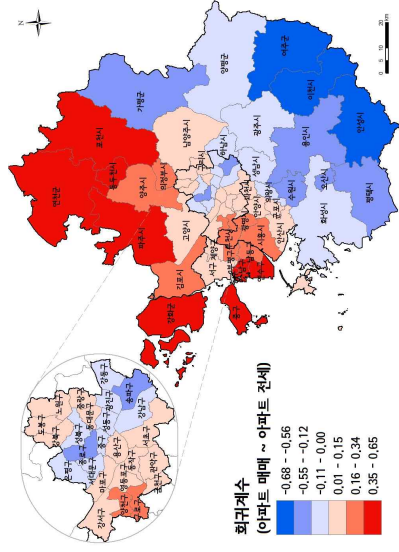
a) 아파트 매매가 회귀계수



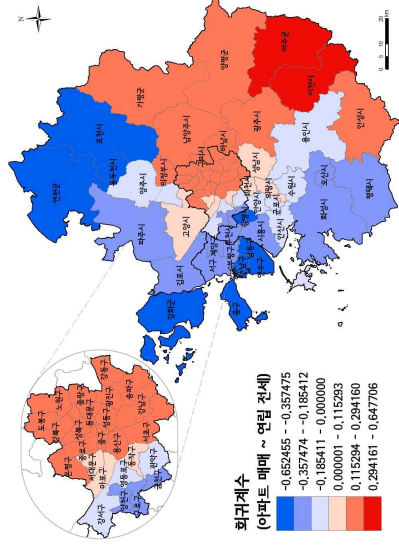
b) 연립 전세가 회귀계수



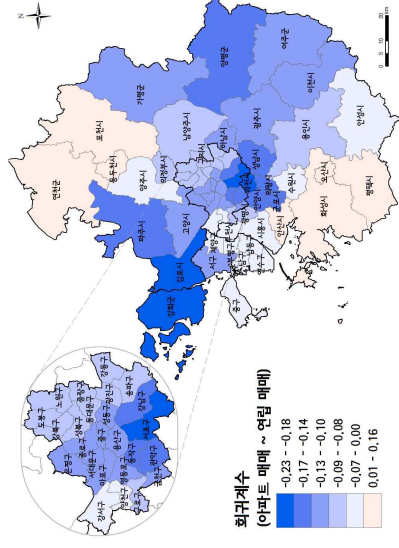
c) 연립 매매가 회귀계수



d) 아파트 전세가 회귀계수



e) 연립 전세가 회귀계수



f) 연립 매매가 회귀계수

[그림 4-4] GWR 회귀계수(종속변수 상: 아파트 전세가, 하: 아파트 매매가)

제 2 절 수도권 주택 가격의 시공간적 상호작용 분석

본 절에서는 SpVAR 모델을 통해 수도권 주택 하위시장 간 가격 모델을 제시하고자 한다. 앞선 절에서 분석한 수도권 주택 가격의 시공간적 특성은 SpVAR 모델을 활용하기에 적합한 것으로 나타났으므로 1차 차분된 주택 가격, 즉 가격 변동량을 대상으로 SpVAR 모델을 수립하고자 한다. 이러한 SpVAR 모델을 기존의 추정방법인 SUR과 본 연구에서 제안한 SpVAR-Lasso를 통해 각각 전역적 모델과 국지적 모델을 도출한 뒤 이들 간의 정확성 비교를 통해 수도권 주택 시장에 보다 적합한 모델을 제시하고자 한다. 분석에 앞서 시차와 공간 가중치에 대한 사전적인 설정이 반영되었다. 시차의 경우 1차에서 3차까지 설정하여 분석한 결과 1차에서 가장 효율적인 결과를 보였다. 또한 공간 행렬(W)은 1차 인접행렬, 2차 인접행렬, 역거리제곱행렬, 역네트워크거리제곱행렬을 적용해본 결과 인접행렬의 결과물이 가장 우수하였다. 그러므로 본 연구에서는 시차가 1이고 공간적 범위 역시 1인 SpVAR(1,1) 모델을 기준으로 분석하고 모델 간 비교를 실시하였다.

1) 주택 매매-전세가격의 전역적 상호작용 패턴

본 연구에서 제시한 수도권 주택의 SpVAR 모델은 아래 [식 4-4]와 같다.

$$Y_t = \Gamma_0 + \Gamma_1 Y_{t-1} + \Gamma_2 Y_{t-1}^* + u_t \quad [\text{식 4-4}]$$

단,

$$Y_t = [\Delta RA_t, \Delta PA_t, \Delta RB_t, \Delta PB_t, \Delta RC_t, \Delta PC_t]^T$$

$$Y_{t-1} = [\Delta RA_{t-1}, \Delta PA_{t-1}, \Delta RB_{t-1}, \Delta PB_{t-1}, \Delta RC_{t-1}, \Delta PC_{t-1}]^T$$

$$Y_{t-1}^* = W^* [\Delta RA_{t-1}, \Delta PA_{t-1}, \Delta RB_{t-1}, \Delta PB_{t-1}, \Delta RC_{t-1}, \Delta PC_{t-1}]^T$$

이 식은 각각 52개($T=52$)의 사례수와 12개(시차변수 6개, 지연된 공간차 변수 6개)의 설명변수가 있는 방정식 390개(65개 지역, 6개 변수)를 포함

하고 있다. 기존의 추정에서는 사례수에 비해 많은 변수의 개수 문제를 해결하고자 각 변수 별로 독립적인 블록으로 통합하였다. 즉, 각각의 변수별 블록에는 65개 지역의 시계열자료가 통합됨으로써 총 3,380개($N=65$, $T=52$)의 사례수를 확보함으로써 블록별 SUR 추정이 가능하게 되었다. 이에 따라 전 지역에서의 변수 간의 관계는 동질적으로 가정되며, 한 지역의 특정 변수에 영향을 미치는 다른 지역의 변수는 전역적인 변수 간 관계로 가정된다. 예를 들어, 당기의 강남구 아파트 전세가는 전기의 서초구의 아파트 매매가와 송파구의 아파트 매매가에 의해 각기 다른 영향을 받을 수 있지만, SUR 추정에서는 이를 수도권 전역에서의 전세가와 매매가 간의 관계로 간주한다.

각 블록별로 추정된 계수는 다음의 <표 4-9>와 같다. 이에 따르면 특정 지역의 아파트 전세의 가격 변화량은 해당 지역의 전기 전세가 변화량과는 반비례(-0.276)하지만 전기의 가중된 전세가 변화량에는 비례(0.221)한다. 이는 직관적인 인식과 부합하는 결과로 이해할 수 있다. 지난달에 전세가가 급등할 경우, 시장에서 이를 비싸다고 여겨 소비자들의 수요가 감소함에 따라 가격 변화량 낮아지는 현상을 보이기 때문이다. 그러나 주변 지역의 영향력이 포함된 가격의 전기 값이 긍정적인 영향을 미친다는 것은 결국 주변지역의 전세가 상승이 해당지역의 전세가 상승에 영향이 있으며, 주택 가격이 주변으로 시차를 두고 확산해 나가고 있음을 뒷받침하는 결과로 볼 수 있다. 또한 아파트 전세가의 변화는 아파트 매매가 변화량에 비례(0.018)한다. 이는 매매가 상승폭이 가파를 경우, 공급자는 전세 물량을 매매로 전환하고자 하며 소비자는 아파트를 구매하는 대신 전세를 선택하게 되기 때문에 전세 물량이 수요에 비해 공급이 부족해지게 됨에 따라 가격이 상승하는 것으로 사료된다. 반면 아파트 매매의 경우, 아파트 전세와의 관계가 유의하지 않아 아파트 전세가 변화가 매매가 변화에 영향을 준다고 보기 어렵다. 이는 임승직(1996)이 정의한 바와 같이 국내 아파트 시장에서 매매가 임대료의 상급재라는 가설을 뒷받침하는 결과이다.

또한 아파트 전세가의 상승폭이 커질 때, 연립다세대 매매가 역시 상승하게 되는데 이는 가격 이외의 요소에서 발생하는 상급재적 요소로 인한

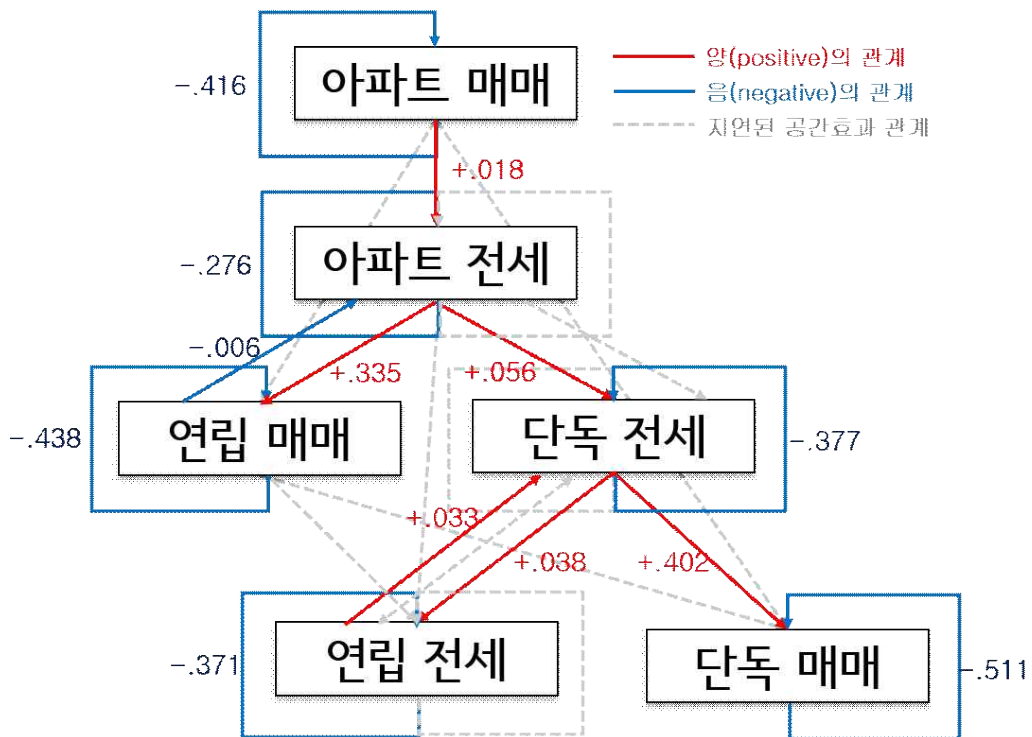
결과로 파악된다. 기간 내 연립다세대의 평균 매매가격은 아파트의 평균 전세가를 상회하여 아파트 전세가 연립다세대 매매의 공급재라고 보기 어렵다. 그러나 우리나라의 주택 인식 상 아파트가 연립주택에 비해 고급재화라는 인식이 있어 소유형태와 무관하게 아파트에 거주하는 것을 선호한다. 이러한 인식으로 인해 아파트 전세가가 빠르게 상승하게 되면, 이에 대한 대체 재화로 연립다세대 주택을 선택하기 때문에 그의 매매가 역시 빠르게 상승할 수 있음을 시사한다.

<표 4-9> SpVAR 추정 계수(SUR)

		전세			매매		
		아파트	연립	단독	아파트	연립	단독
상수항		1.874 ***	0.700 **	0.401 **	0.274	-0.344	-0.484
시차 변수	RA	-0.276 ***	0.034	0.056 ***	-0.035	0.335 ***	0.282
	PA	0.018 **	0.004	-0.002	-0.416 ***	-0.071 *	-0.084
	RB	0.000	-0.371 ***	0.033 **	0.019	0.063	-0.164
	PB	-0.006 **	0.005	0.000	-0.001	-0.438 ***	0.023
	RC	0.011	0.038 *	-0.377 ***	0.056 *	0.106	0.402 **
	PC	0.000	0.000	0.000	-0.001	0.002	-0.511 ***
지연된 공간차 변수	RA	0.221 ***	0.286 ***	0.222 ***	0.017	-0.081	0.275
	PA	0.004	-0.043 *	0.037 *	0.016	0.175 **	-0.590 *
	RB	0.007	0.159 ***	0.101 ***	0.080	-0.095	-0.066
	PB	0.009	0.031 **	0.008	0.012	0.010	-0.313 **
	RC	0.039	0.180 ***	0.167 ***	0.060	-0.087	0.137
	PC	-0.002	0.000	-0.001	0.002	-0.008	-0.025
Adj R ²		0.075	0.145	0.152	0.166	0.199	0.266

***: 99%, **: 95%, *: 90% 신뢰구간에서 유의함

이를 보다 시각적으로 이해하기 쉽게 표현한 모델이 [그림 4-5]이다. 이를 살펴보면, 아파트 매매가 가장 상위 재화로 위치하며 자기회귀적인 영향만 받을 뿐 다른 변수에 의해 설명되지 않는다. 결국 아파트 매매 가격은 다른 주택 종류의 가격에 대해 독립적으로 결정되며, 부적 피드백 작용을 받으며 상승폭이 제한된다. 이렇게 결정된 아파트 매매 가격은 아파트 전세시장에 긍정정인 영향을 준다. 비록 큰 영향력은 아니지만, 매매 가격의 증가 폭이 상승할 때, 아파트 전세 가격 상승률도 높아지는 것을 확인할 수 있다. 이는 최근 주택 시장에 대해 전세 가격 변화가 시장을 선도한다는 언론이나 대중적 믿음과 배치되는 결과라 할 수 있다. 대신, 높아진 아파트 전세가를 대체하기 위한 수단으로 연립주택을 매매하거나 단독주택을 전세로 구하는 경우가 발생함을 확인할 수 있다. 이 때, 두 대체 재화 중 보다 큰 영향을 받는 연립주택 매매의 대체효과가 보다 크다고 할 수 있다.



[그림 4-5] 전역적 SpVAR 모델의 추정 결과

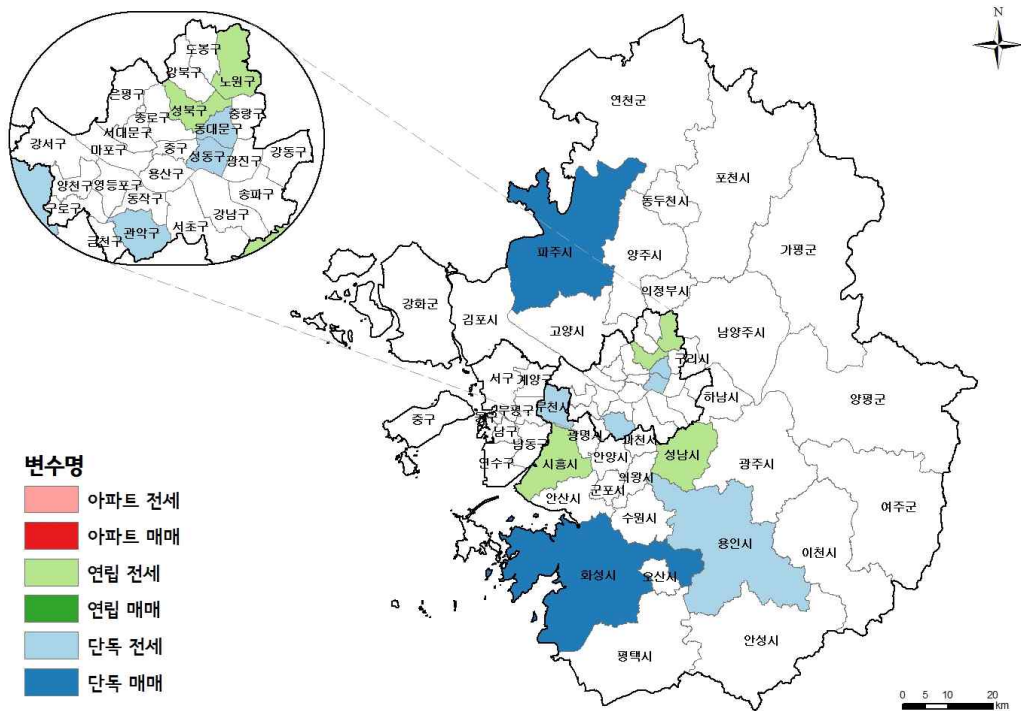
2) 주택 매매-전세가격의 시공간적 대체효과

앞서 SUR로 추정한 SpVAR 모델은 수도권 주택시장의 등질성을 가정으로 한 전역적인 관계를 파악할 수 있는 결과를 도출하였다. 그러나 이는 국지적 수준에서 이질적인 패턴을 보기 어려운 한계를 갖는다. 특히 국내 주택시장에서는 가격차로 인한 공간적 위계성이 발생하므로 이와 같은 전역적 결과가 구체적인 지역에서의 주택 가격 관계를 설명하기 어렵다. 이를 극복하기 위한 방법으로 제시한 SpVAR-Lasso는 앞선 [식 4-4]에서 발생하는 자유도 문제와 과소표본의 문제를 Lasso 추정을 통해 해결함으로써 변수별로 자료를 통합으로 발생하는 정보의 손실을 막을 수 있다. 즉, SUR에서는 각각의 방정식에 해당 지역 변수만 포함하여 12개의 변수만을 가졌다면, Lasso에서는 780개의 변수를 모두 방정식에 투입하여 연산이 가능하다. 이에 따라 다른 지역의 다른 변수에 의한 영향력을 확인할 수 있다.

추정 결과 총 304,200개 회귀계수 중 6,084개가 유의하게 도출되었다. 이 중 1절에서 가장 높은 가격으로 나타난 지역의 가격 변동이 주변지역에 미치는 영향을 확인하고자 한다. 우선 강남구 아파트의 전세가 상승은 다른 하위 시장의 영향을 받지 않았으며, 어느 지역의 아파트 매매가 변화에 영향을 주지 못하였다. 반면 강남구의 아파트 매매가는 강남구 아파트 전세에 대한 영향이 없는 것으로 나타나 전역적인 수준에서 확인되었던 아파트 매매가 상급재인 대체 관계가 강남구에서는 성립되지 않음을 확인하였다. 이는 강남구의 전세 시장은 매매 시장과는 독립되어 있으며, 다른 하위시장과 연관 되어 있음을 의미한다.

또한 강남구의 전세가는 자기 자신의 과거값에 의한 영향이 발생하지 않아, 외부적 조건이 갖추어진다면 과거의 상승과는 무관하게 가격이 상승하는 구조를 갖고 있다. 즉, 강남구 아파트의 사용가치가 주변 환경 등의 개선으로 높아지기만 한다면, 이전의 전세가와는 무관하게 계속 상승할 가능성이 있음을 시사한다. 한편, 강남구의 전세가 상승은 주변지역의 아파트보다는 다른 형태의 주거에 영향을 주는 것으로 나타났는데, 서울 외곽

또는 서울 근교로 그 영향력이 전해지고 있다. 연립다세대 주택의 전세가는 성남시, 시흥시, 노원구, 성북구에서 정적으로 반응하며, 단독다가구 주택은 보다 더 먼 지역인 부천시, 인천 남구까지 영향을 받는 것으로 나타났다(그림 4-6).



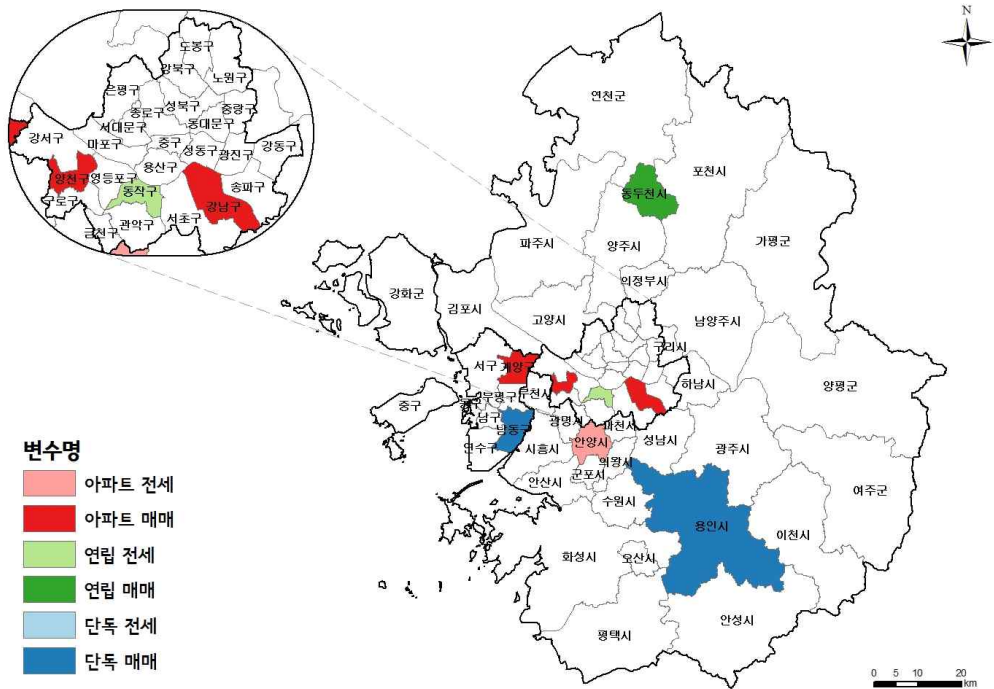
[그림 4-6] 강남구 아파트 전세가에 의해 영향을 받는 지역

반면, 강남의 아파트 매매가는 아파트 전세 가격에 비해 주변 지역과의 상호작용이 활발하게 나타났다. 우선 강남구 아파트 매매가는 전기 가격에 의한 피드백 작용이 발생하여, 높아지는 매매가격에 제약이 가해지는 것으로 나타났다. 이는 강남구 아파트 전세에서 보이지 않는 작용으로, 2011년 이후 전세가의 상승에 비해 아파트 매매가격이 안정적인 원인으로 판단된다. 또한 강남 외의 아파트 밀집지역으로 볼 수 있는 양천구의 아파트 매매가격의 영향도 받는 것으로 나타났으며, 이 역시도 부적 관계를 보였다. 이는 강남과 목동의 아파트 매매 시장이 상호 피드백 관계에 있어 어

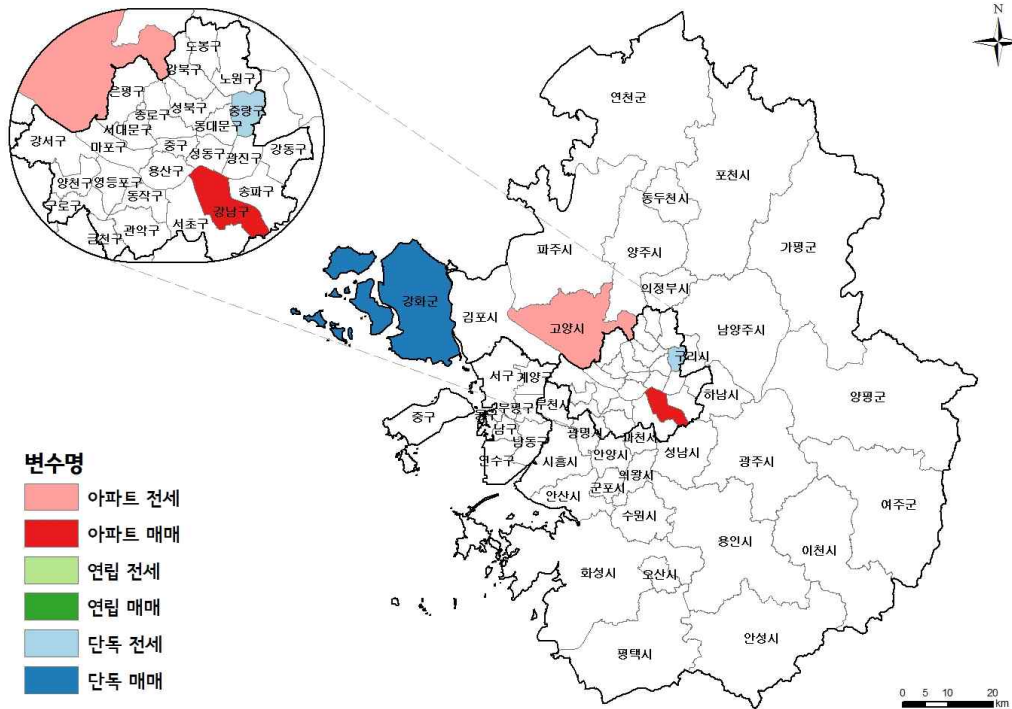
느 정도 상승폭을 제한하는 기재로 작용하고 있다고 판단된다. 뿐만 아니라 서울 동작구와 경기 남부권의 주택 시장이 강남 아파트 매매시장에 영향을 주기도 하는 현상 역시 발견되었다(그림 4-7).

한편 강남 아파트 매매가에 긍정적인 영향을 받는 하위 시장은 고양시의 아파트 전세 시장과 중랑구의 단독 전세 시장으로 나타났다. 이는 강남 아파트 매매의 대체효과를 보여주며, 이는 강남의 아파트 가격이 고양시의 아파트 가격과 밀접한 관계가 있다는 김경민(2007)의 선행연구와 일관적인 결과를 보이고 있다. 그러나 강화군의 단독 매매 시장과는 부적 관계가 도출되었는데, 이는 이례적인 현상으로 판단되며 강화군의 적은 거래 건수로 인한 문제로 보인다(그림 4-8).

그러나 모델에서 추정된 계수의 값을 해석하는데 있어 변수 간 내생성으로 인해 그 수치를 정확히 이해하기 어렵다. 이러한 문제를 해결하기 위해 3장에서 제시한 충격반응분석을 적용하였다. 대신 SUR 추정과 달리 이질성을 반영하였으므로 오차항에서의 공간 관계를 따로 정의하지 않고 오차항의 구조를 그대로 활용할 수 있다. 이를 통해 지역의 주택 가격이 다른 지역의 하위시장 가격에 미치는 영향을 정량적으로 분석하였다. 이때, 주택 가격이 가장 높은 강남구를 중심으로 이 지역에 외부적 충격으로 인한 변화가 다른 지역에 어떠한 변화를 가져오는지 살펴보았다. 여기서 외부적 충격이란 모델에 포함된 변수(각 지역별, 주택 종류별 거래 가격의 변화량) 이외의 요소에 의해 해당 종속변수가 변하는 것을 의미한다. 예를 들어, 개발 호재에 의해 강남구의 아파트 전세가가 오르게 되는 경우를 의미한다. 이러한 외부 충격이 강남구 아파트 전세가의 오차항을 1단위 증가시키게 되며, 이 변화는 국지적 SpVAR로 추정된 수도권 주택 하위시장의 매매와 전세가격 시스템에 의해 다른 하위주택시장에 영향을 미치게 된다. 즉, 충격반응량은 강남구의 전세가의 변화가 1만원/㎡만큼 오를 경우, 해당 하위시장의 어느 정도 오르는지를 보여주는 지표를 말한다.



[그림 4-7] 강남구 아파트 매매가에 영향을 주는 지역



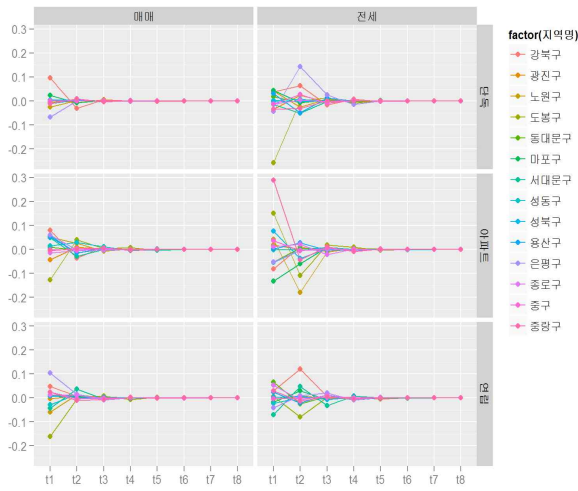
[그림 4-8] 강남구 아파트 매매가에 의해 영향을 받는 지역

<표 4-10>은 강남구 아파트 전세가에 충격이 발생했을 경우 수도권 지역의 주택 하부시장에서 발생한 4기까지의 충격반응량을 합한 것이다. 이를 살펴보면, 전역적인 수준에서 아파트 전세 가격은 즉시 크게 반응하며 연립다세대의 매매가격이 그 뒤를 따른다. 반면, 아파트 매매가격과 연립다세대 전세가격은 그 즉각적인 효과가 미미하거나 음의 방향으로 나타나지만 시차를 두고 가장 크게 반응한다. 이를 전역적 모델(그림 4-5)에 비추어 보면, 자기회귀효과가 있는 아파트 전세가격과 직접적인 대체 관계가 있는 연립다세대 매매가격은 충격이 바로 나타나는 반면, 직접적인 상관관계가 없는 아파트 매매와 연립다세대의 전세는 다른 경로를 통해 그 효과가 전해지므로 시차를 두고 충격이 나타나는 것으로 해석할 수 있다.

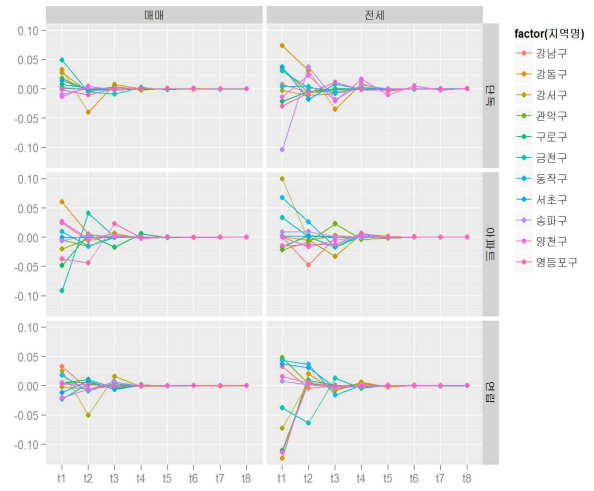
<표 4-10> 강남구 아파트 전세가 변화량에 대한 충격반응량(만원/㎡)

		RA	PA	RB	PB
1기	합계	1.950	0.014	-0.920	0.201
	평균	0.030	0.000	-0.014	0.003
2기	합계	-0.478	0.407	0.356	0.087
	평균	-0.007	0.006	0.005	0.001
3기	합계	-0.006	0.038	0.031	0.005
	평균	-0.000	0.000	0.000	0.000
4기	합계	0.026	0.021	-0.043	-0.008
	평균	0.000	0.000	-0.000	-0.000

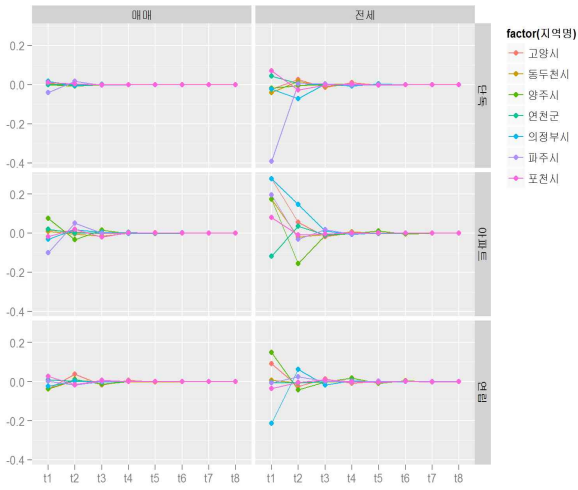
이러한 충격효과가 지역에 따라 차등적으로 전개된다. [그림 4-9]는 강남구 아파트 전세가의 변화에 따른 지역별 충격반응량을 보여주는 것으로 크게 6개 지역으로 구분하여 그 변화를 살펴보았다. 그래프의 열린 거래 유형, 행은 주택 유형으로서 대체적으로 매매에 비해 전세가의 반응이 크게 나타나며 4기를 기점으로 대부분 소멸하는 것으로 나타났다. 예외적으로



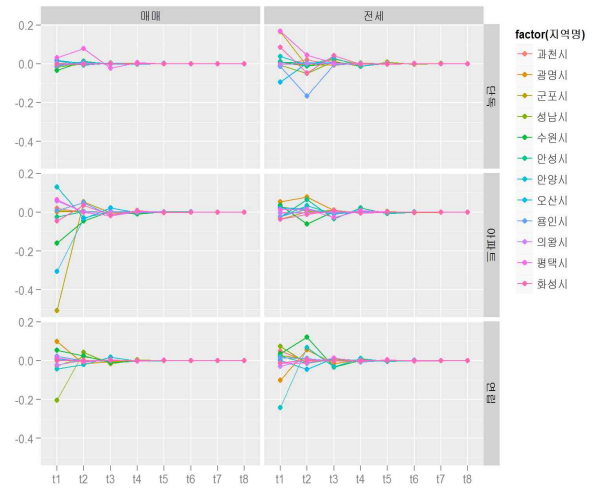
a) 서울 강북



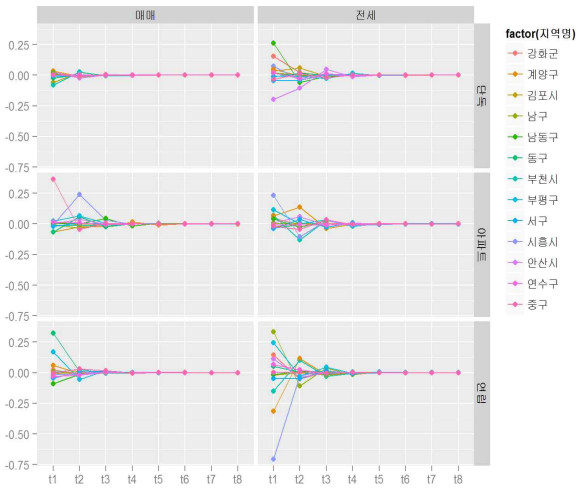
b) 서울 강남



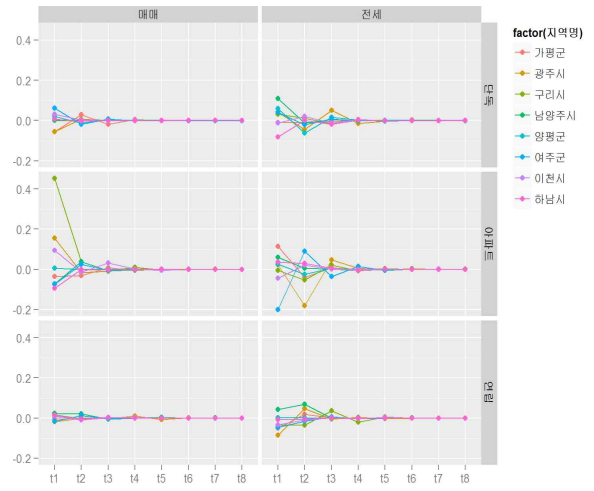
c) 경기 북부



d) 경기 남부



e) 경기 서부



f) 경기 동부

[그림 4-9] 강남구 아파트 전세가 변화에 의한 지역별 충격반응량

구리시의 아파트 매매가가 강남 아파트 전세가 상승에 크게 동조효과를 보이고 있으며(그림 4-9f), 서울 강북권의 아파트 전세 역시 다른 지역에 비해 강남구 아파트 전세가 변화에 민감하게 반응한다(그림 4-9a).

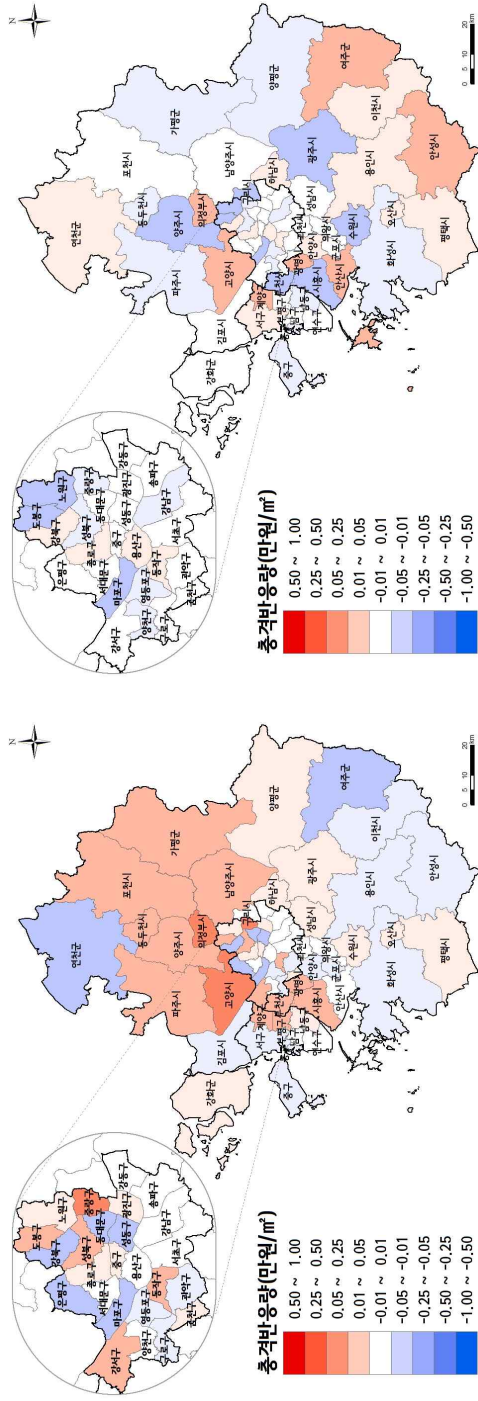
이를 공간적으로 살펴보기 위해 지도로 표현하였다(그림 4-10~13). 충격이 가해진 후 1기부터 4기까지의 충격량의 변화가 그림 좌상단-우상단-좌하단-우하단 순으로 전개되어 있다. 우선 아파트 전세가의 변화량을 살펴보면(그림 4-10), 오히려 접경지역보다는 조금 떨어진 경기 북부와 서부 지역에 가장 큰 영향이 미치고 있다. 특히 일산으로 대표할 수 있는 고양시의 변화량(0.28)이 두드러지는 것을 확인할 수 있으며, 중랑구(0.29)와 의정부(0.28)의 변화도 크게 나타났다. 이러한 변화는 시간이 흘러감에 따라 완화되는 것을 확인할 수 있는데, 완화되는 시기가 지역마다 차이를 보인다. 중랑구의 경우 그 바로 다음 시기에 감소세로 전환되어, 전기의 상승폭을 둔화시키는 작용이 발생하는 반면, 고양시의 경우 2기까지 의정부시의 경우 3기까지 누적적으로 발생한다. 충격 이후 2기에서는 1기에 비해 원거리 지역의 전세가가 상승하는데 주로 경기 남부권에서의 변화가 눈에 띈다. 이는 강남구 전세가 상승은 시차를 두고 퍼져나가는 현상을 잘 보여준다. 3기에서는 피드백 작용이 나타남으로써 2기와 반대의 패턴이 나타난다. 예를 들어 2기에서 하락을 보인 광주시의 경우 3기에서 충격량이 커지는 반면 안성시의 경우 반대의 패턴을 보인다.

한편 강남구 아파트의 전세가는 주변 지역의 아파트 매매가에도 영향을 준다. [그림 4-11]을 보면 수도권 전역에서 비교적 큰 충격을 받은 아파트 전세가와 달리 아파트 매매가는 크게 북부 축, 동부 축, 남부 축을 중심으로 두드러지는 패턴이 나타난다. 북부 축은 강남구에서 시작하여 성동구(0.014) - 용산구(0.05) - 성북구(0.06) - 강북구(0.08) - 양주시(0.08)로 이어지며, 동부 축은 광주시(0.15)와 이천시(0.09)의 충격량이 크게 나타났다. 남부 축은 과천시(0.02), 안양시(0.13), 의왕시(0.06)로 이뤄진 서울 근교의 신도시 밀집 지역에서 높게 양의 반응을 보인 반면, 이 보다 조금 떨어진 군포시, 수원시, 화성시, 오산시에서는 음의 충격량을 보였다. 이러한 1기의 충격량은 2기에서 축을 중심으로 주변지역에 영향을 미치는 것을

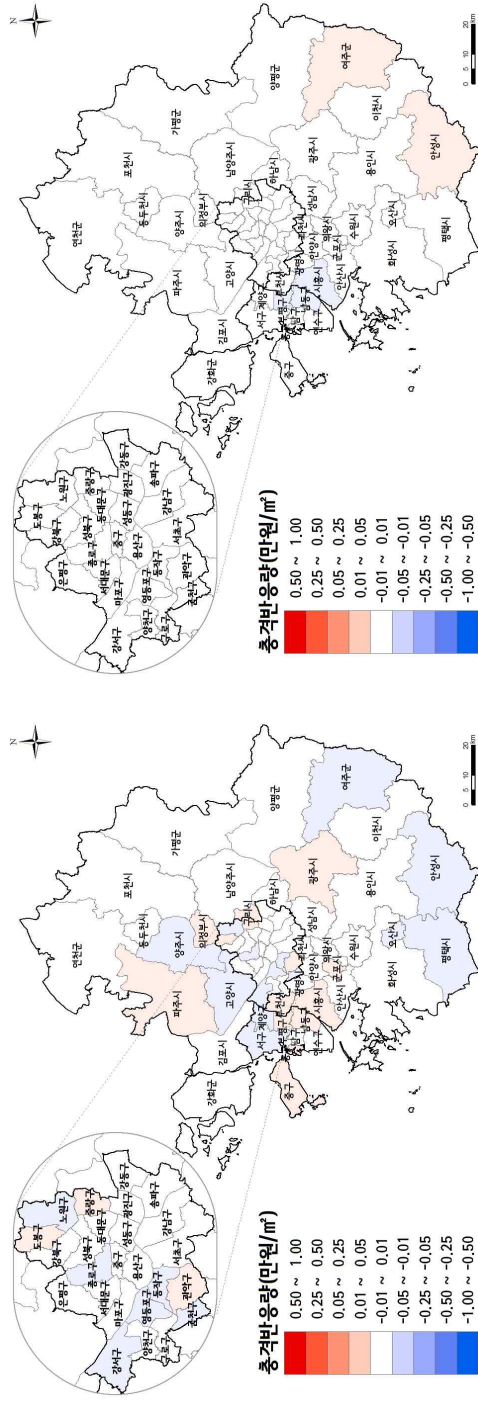
볼 수 있다. 양주시와 인접한 파주시, 포천시의 경우 1기의 음의 충격량이 양의 충격량으로 변화하였으며, 광주시, 이천시와 인접한 여주군의 경우도 같은 패턴을 보였다. 경기 남부지역에서는 안양시와 의왕시에 인접한 군포시가 양의 충격량으로 변환되었으며, 시흥시와 안산시까지 그 영향력이 미치기 시작하였다. 이는 강남 아파트 전세가격의 상승이 시차를 두고 공간적으로 확산되어 수도권 아파트 매매가격에 영향을 미치고 있음을 보여준다. 그러나 이러한 반응은 아파트 전세가와 같이 피드백 과정을 거쳐 점차 소멸되어, 강남 아파트 전세가의 상승이 지속적이지 않음을 알 수 있다.

[그림 4-12]는 강남구 아파트 전세가 변화에 따른 수도권 연립다세대 주택 전세가의 변화로서 아파트의 전세거나 매매가가 받는 영향력에 비해 그 충격량이 작아 경기도 최외곽까지 영향력이 미치지 않는다. 대신 아파트에 비해 인접한 지역에서의 양의 충격반응이 비교적 크게 나타났다. 강남구를 인근의 관악구(0.05), 서초구(0.04), 동작구(0.04), 용산구(0.02), 영등포구(0.02)에서 1기에 바로 반응이 나타나는데, 이는 아파트 가격에서 보인 패턴과는 사뭇 다르다. 이는 주택의 사용가치에 근거한 전세 수요가 해당 지역의 전세 가격이 올랐을 경우, 생활 환경을 대체할 수 있는 지역을 찾는 것이라고 판단된다. 즉, 강남 지역에 대한 수요가 강남의 어메니티를 활용할 수 있지만 비교적 저렴한 하위시장으로 이전함으로써 해당 지역의 주택 수요가 증가하여 가격 상승을 초래하는 것이다. 아파트의 경우 대부분의 지역에서 높은 가격에 형성되어 있지만, 상대적으로 연립다세대 주택의 가격은 저렴하므로 이러한 공간적 대체효과가 아파트가 아닌 연립다세대 시장에서 발견되는 것으로 보인다.

마지막으로 [그림 4-13]은 연립다세대 매매시장의 충격반응으로서 인천 지역(계양구, 부평구, 동구)의 변화가 크게 나타나는 편이다. 이는 강남의 전세가 상승으로 인해 발생한 수요의 이전이 비교적 인접한 지역으로 가더라도 매매 보다 저렴한 전세의 특성상 인근 지역의 전세시장으로 유입되는 반면, 연립다세대 매매시장의 경우 접근성이 더욱 떨어지는 지역까지 밀려나는 현상을 보이고 있다. 이러한 일련의 변화들은 강남구의 아파트 전세가격에 의해 수도권 주택별 하위 시장이 반응함을 입증하고 있다.

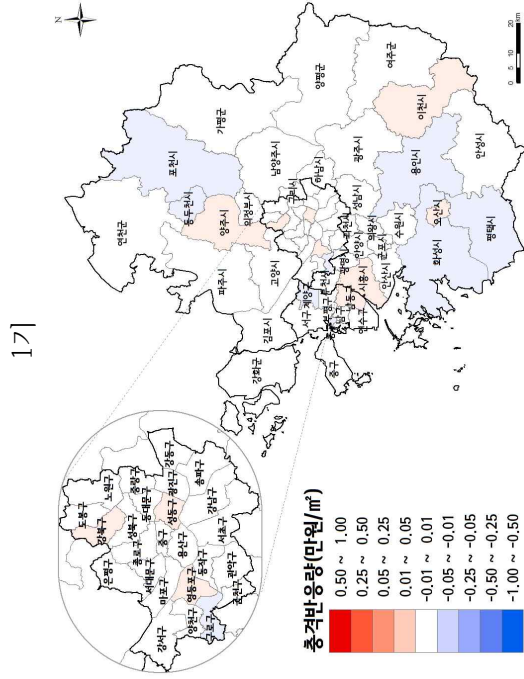
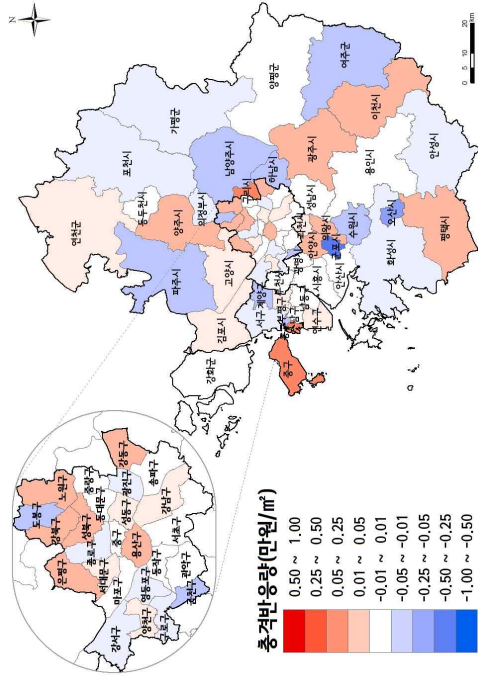
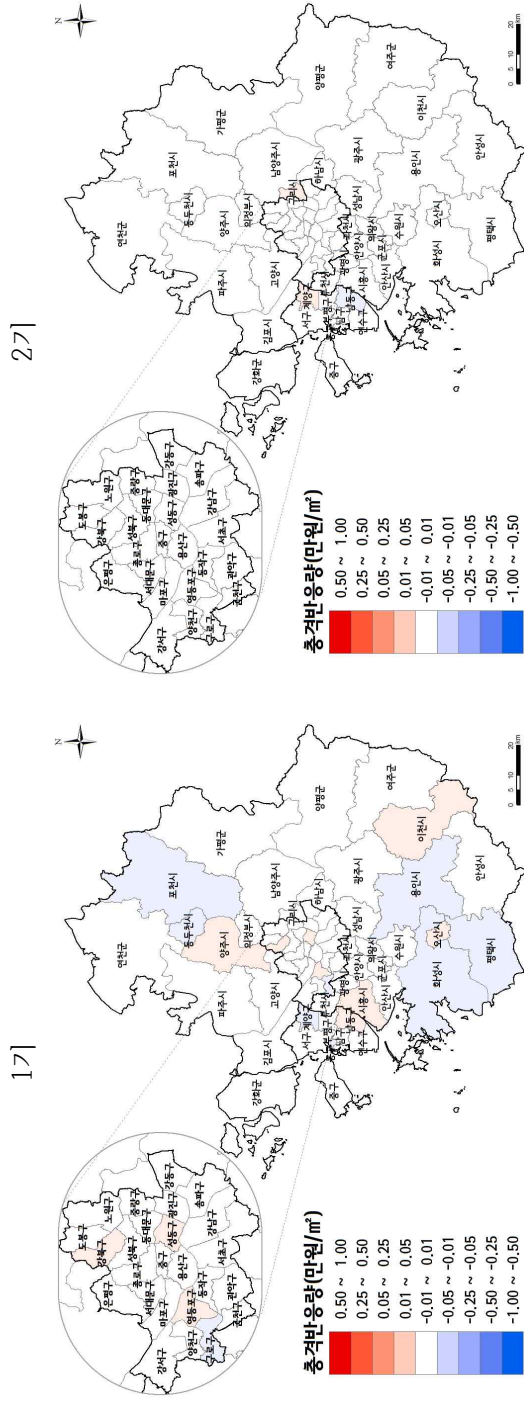
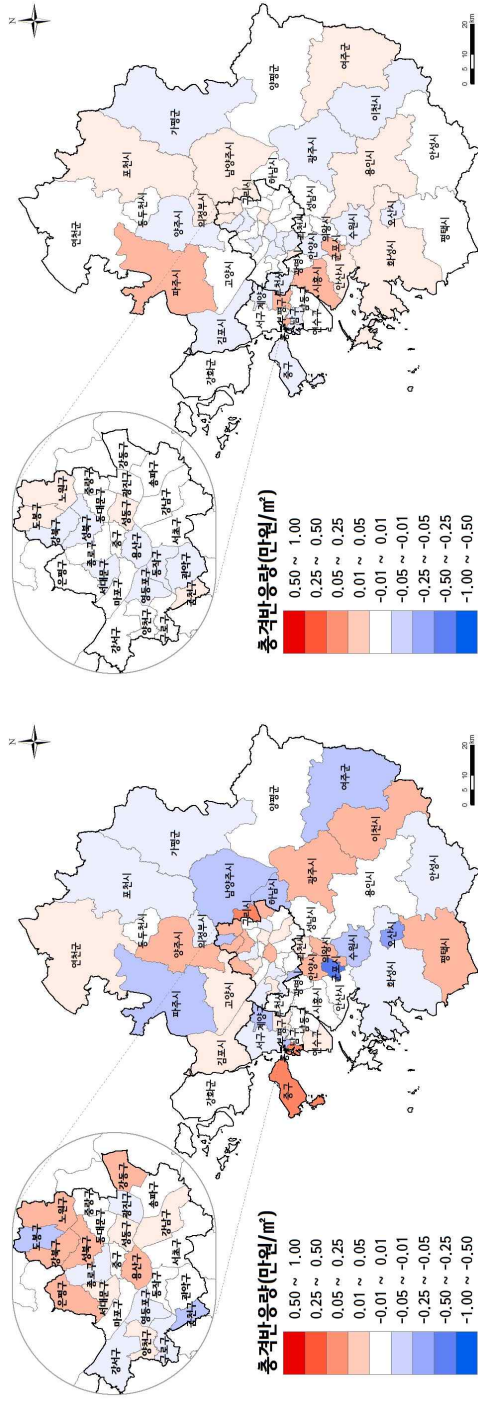


17]

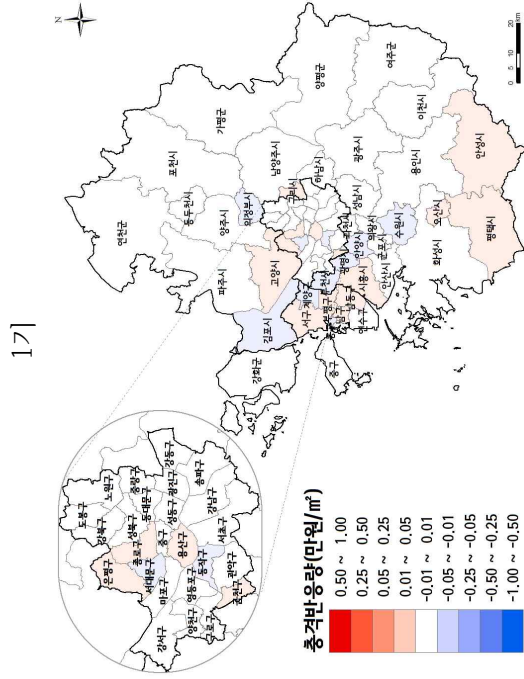
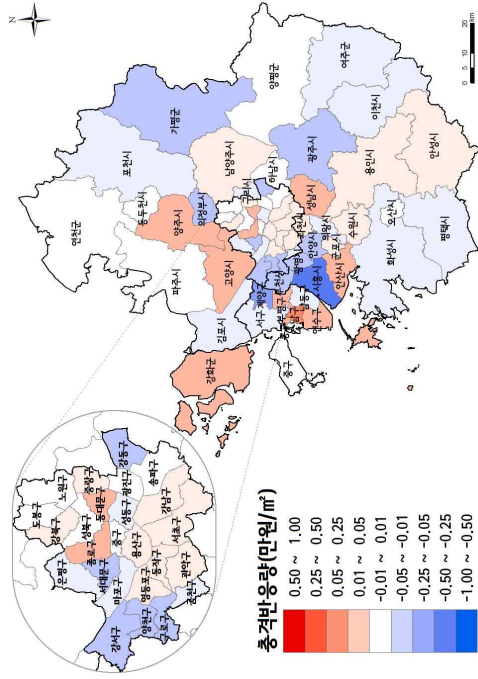
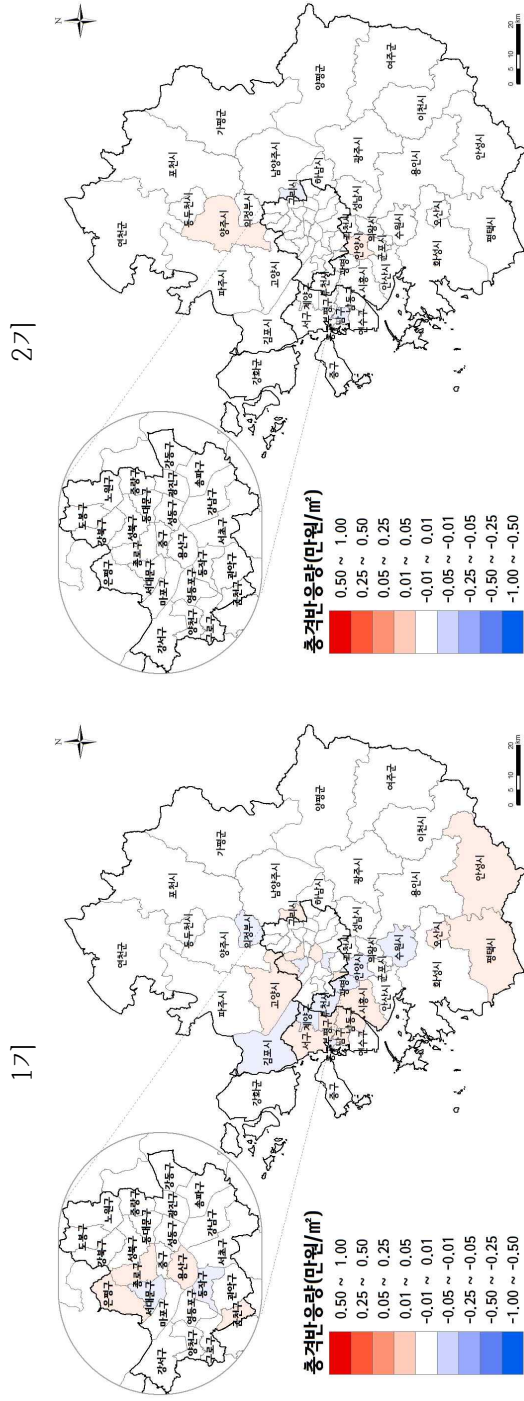
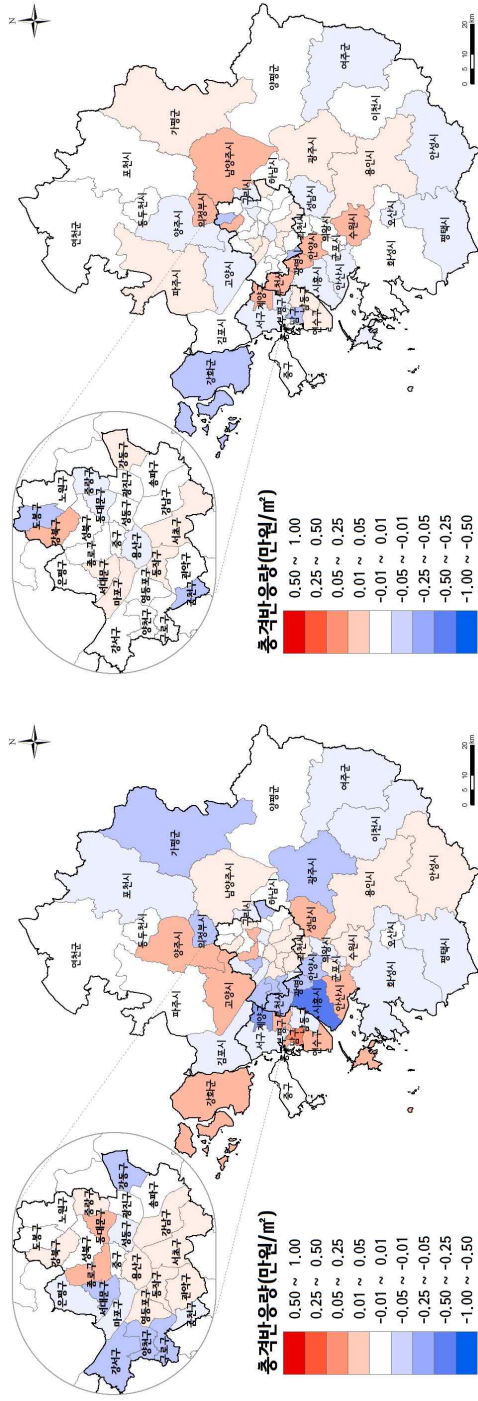


37]

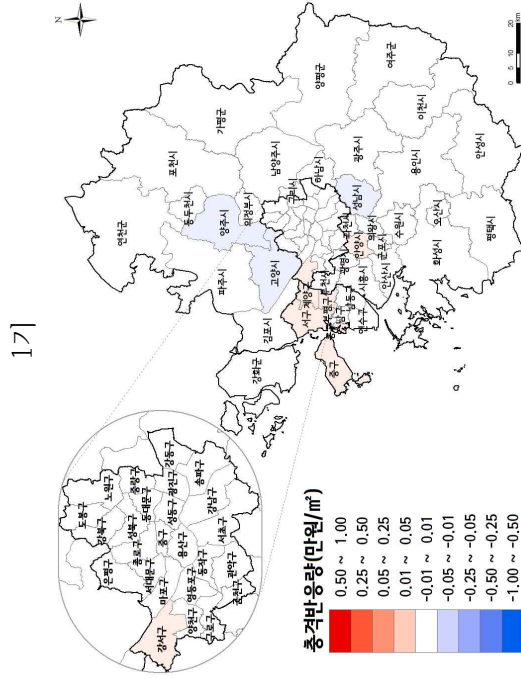
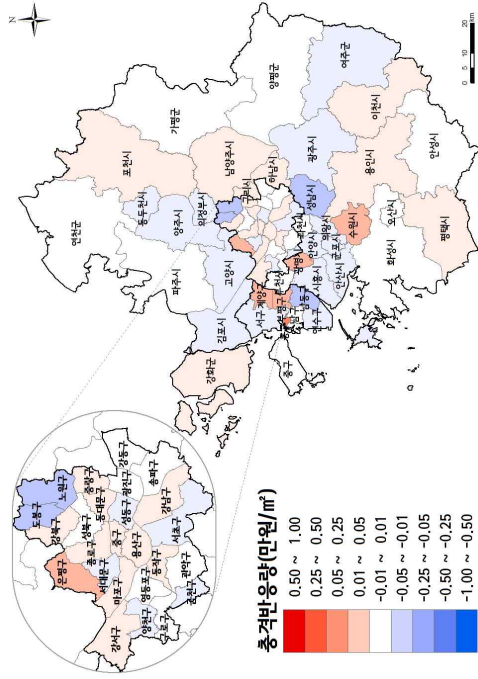
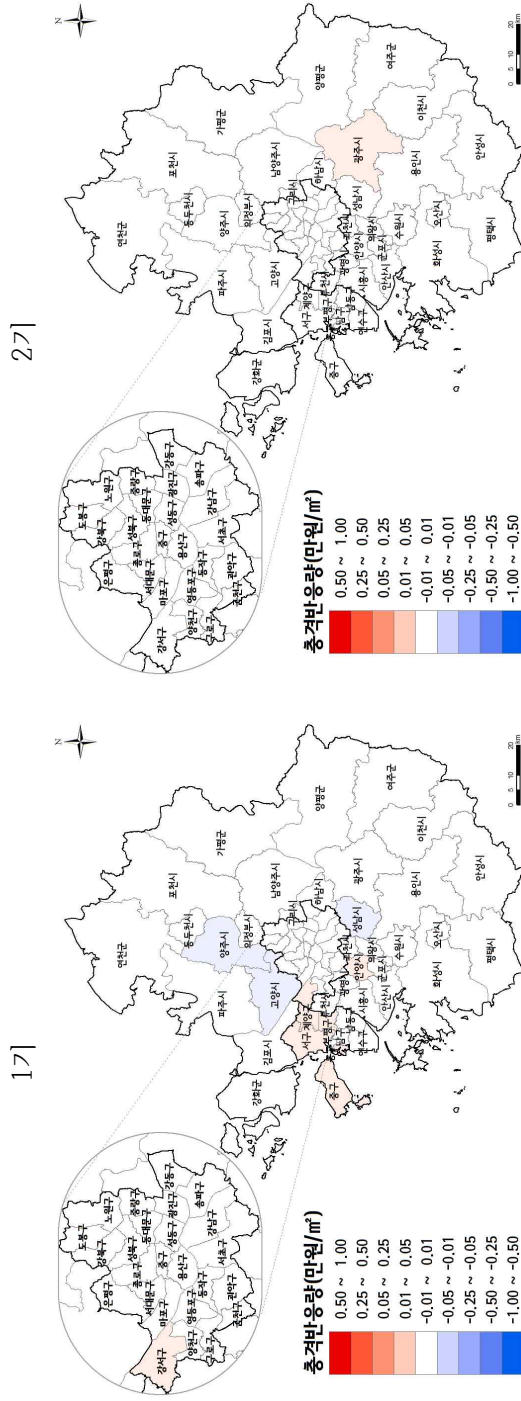
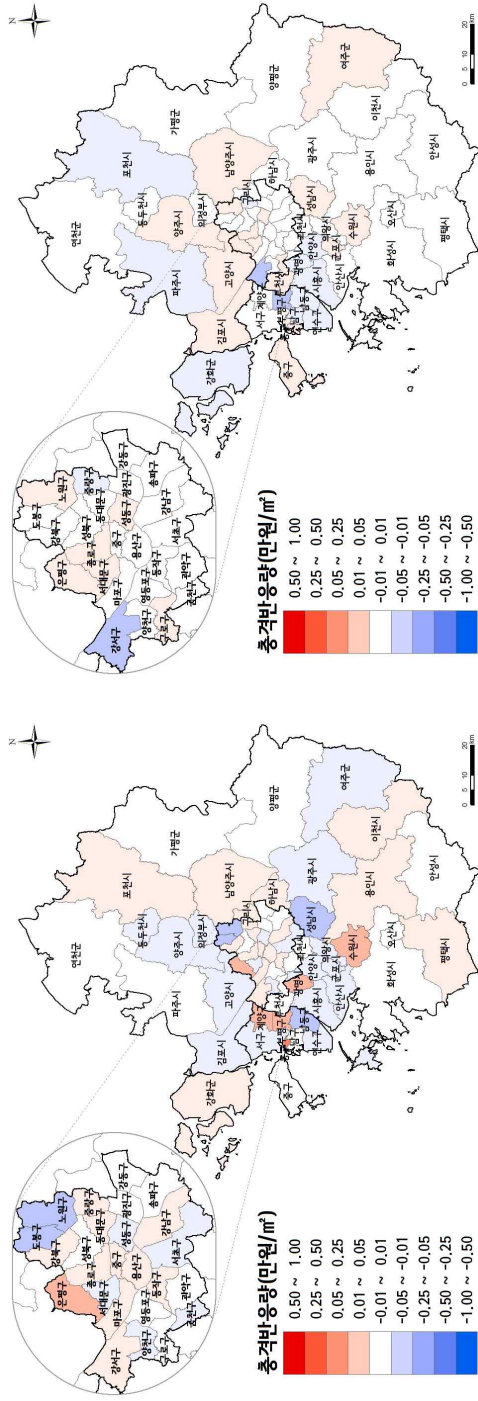
[그림 4-10] 강남구 아파트 전세가 변화량에 대한 수도권 아파트 전세가의 충격반응량



[그림 4-11] 강남구 아파트 전세가 변화량에 대한 수도권 아파트 매매가의 충격반응량



[그림 4-12] 강남구 아파트 전세가 변화량에 대한 수도권 연립다세대 전세가의 충격반응량



[그림 4-13] 강남구 아파트 전세가 변화량에 대한 수도권 연립다세대 매매가의 충격반응량

그러나 이러한 강남구 아파트 전세가 변화에 의한 하위시장 가격의 변화는 크지 않은 것으로 나타났다. 이는 다양한 변수들이 개입하는 주택 가격에 대하여 다른 지역의 가격으로만 설명하고자 했다는 근본적인 한계에서 비롯된다고 볼 수 있다. 하지만 분산분해는 충격반응의 효과를 보다 면밀히 살펴보기 위한 방법으로서 수도권 주택 하위시장의 가격 간 존재하는 상호작용을 쌍방으로 확인할 수 있다. 본 연구에서는 한 단위 충격으로부터 발생한 이후 4차까지의 오차에 대한 설명비율을 살펴보았다. 이는 대부분의 변수에서 4차 이후 충격반응량이 0으로 수렴하기 때문이다.

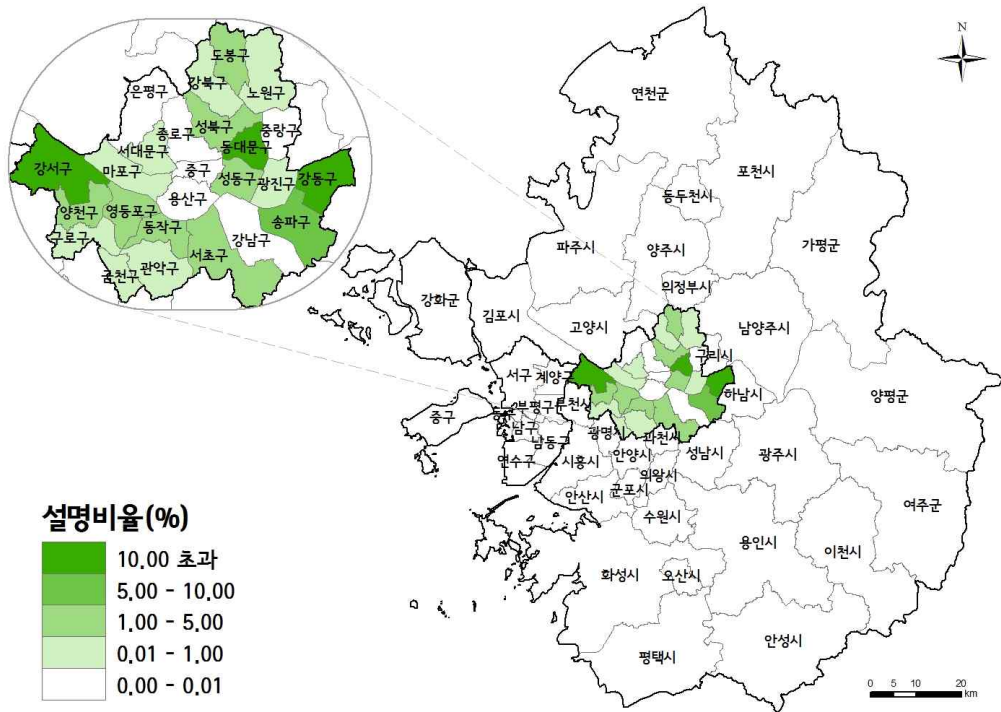
우선 강남구의 전세가 변화는 주로 서울의 전세가 변화를 크게 설명한다. [그림 4-14a]는 강남구 아파트 전세가격의 변화가 해당 지역에서 발생한 변화를 설명하는 비율을 나타낸 지도이다. 이를 보면 강서구 전세가의 변화는 약 17%정도로 가장 크게 설명된다. 뒤이어 동대문구(15.8%), 강동구(11.8%) 역시 상당히 높게 설명된다. 또한 인접한 송파구와 서초구의 전세가도 일부분 강남구의 전세가로 설명되는 부분이 있다. 가장 높은 설명비율을 보인 강서구를 보다 자세히 살펴보면, 앞서 1항의 충격반응분석에서 1기에 약 0.1의 충격반응량을 보였다. 이는 강남구 아파트 전세가에 의한 강서구 아파트 전세가의 0.1만원/㎡만큼의 변동이 강서구의 아파트 전세가격의 전체 변화량의 17%를 차지함을 의미한다. 이는 최근 강서구까지 연결된 9호선의 영향으로 인해 강남과의 접근성이 높아져 강남구의 어메니티를 활용할 수 있는 대체 시장으로 자리매김하고 있는 것이 아닐지 유추하는 근거가 된다.

그러나 강남구의 전세가는 아파트 매매가의 변화에 대한 설명은 크게 하지 못한다. 가장 높은 비중을 차지하는 지역이 성북구이지만 약 0.05% 정도만을 강남구의 전세가가 설명한다. 즉, 성북구 아파트 매매가의 변동은 강남구의 전세가에 영향을 받고는 있지만 그 이외의 영향에 의해 가격 변동이 크게 나타나는 것으로 볼 수 있다. 또한 다른 종류의 주택 거래 변동을 강남구의 전세가는 크게 좌우하지 않는다. 성남시의 연립다세대 전세가가 가장 크게 설명되지만 0.02% 수준이며, 서울 중구 연립다세대 매매가의 약 0.2%정도를 설명한다. 한편, 두 번째로 높은 전세가를 보이는 서

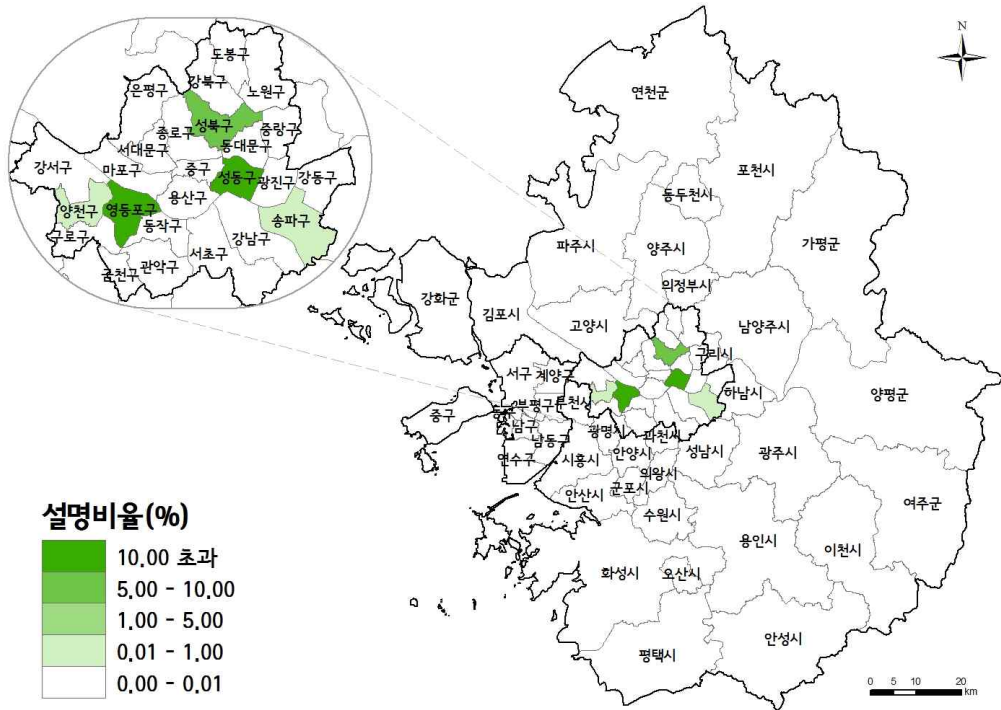
초구의 경우 역시 대부분 지역에서의 아파트 전세가 변화를 설명하지 못한다. 예외적으로 서초구의 전세가가 영등포구와 성동구의 아파트 전세가의 각각 53%, 12% 정도를 설명하고 있지만 다수의 지역에서는 0.01% 이하의 설명력을 갖는다(그림 4-14b).

이상의 국지적 SpVAR 모델을 활용하여 수도권 주택 가격 간 상호작용 패턴을 분석한 내용을 정리하면 다음과 같다. 충격반응분석을 통해 알아본 강남구 아파트 전세가의 상승에 의한 수도권 지역의 주택 하위시장에 미치는 영향은 평균적으로 아파트 전세가와 연립다세대 매매가에 1차적으로 상승효과를 발생시키고, 아파트 매매가와 연립다세대 전세가에 2차적인 상승을 야기함을 확인할 수 있었다. 이는 전역적 SpVAR 모델과 일맥상통하는 결과이다. 이러한 가격 변동을 지역적으로 살펴보면, 공간적으로 상승효과가 확산되는 패턴이 나타난다. 특히 아파트 전세가와 매매가는 시차를 두고 경기 외곽까지 영향을 미침을 확인할 수 있으며, 이러한 영향은 조정작용을 통해 비교적 빠르게 소멸됨을 알 수 있다. 또한 연립주택의 경우, 비교적 가격이 저렴한 전세는 인근지역 혹은 교통 접근성이 탁월한 지역으로 빠른 시기에 그 상승효과가 전달되지만, 매매의 경우 인천 지역 등 비교적 원거리로 강남 아파트의 전세 수요가 이전되는 것을 확인할 수 있었다. 이를 통해 강남구 아파트 전세가 상승에 의한 대체효과가 지가가 비싼 주변 지역에서는 비교적 저렴한 재화로 대체되는 반면 재화의 가격이 상승할수록 대체 지역은 멀어지는 것으로 나타났다.

한편 분산분해 결과 특정 지역의 주택 가격에서 강남 일대의 주택 가격 변화가 설명하는 부분은 극히 일부분으로 나타났다. 예외적으로 서울 외곽인 강서구와 강동구의 아파트 전세가 변화량은 강남구의 아파트 전세가에 의해 10%이상이 설명되었으며, 영등포구와 성동구는 각각 서초구와 송파구에 의해 30% 이상이 설명되었다. 특히 강서구 아파트 전세 시장에서의 변화는 충격반응분석과 분산분해 결과를 종합해보면 강남구 아파트 전세 시장과 밀접한 연관성이 나타난다. 본 연구에서는 이를 교통편의 개선에 따른 접근성의 향상으로 인한 새로운 대체재의 등장으로 추측하고 있으나, 이를 실증적으로 규명하기 위해서는 추가적인 연구가 필요하다.



a) 강남구 전세가의 설명비율



b) 서초구 전세가의 설명비율

[그림 4-14] 지역별 아파트 전세가 변화에 대한 설명 비율

3) 주택 가격의 시공간 상호작용에 대한 SpVAR-Lasso 적용가능성

지금까지 SpVAR 모델들을 활용하여 수도권 주택 가격 간의 시공간적 상호작용을 전역적, 국지적 수준에서 살펴보았다. 기존의 추정방법인 SUR을 활용하여 시공간적 효과가 필터링 된 전역적 SpVAR 모델을 도출하였으며, SpVAR-Lasso를 통해 공간적 이질성을 반영한 국지적 SpVAR 모델을 구축하였다. 각각의 모델은 수도권 주택 시장에서 보이는 상호작용 패턴을 설명할 수 있는 틀을 제시하였다. 전역적 모델을 통해 수도권 지역에서 여전히 아파트 매매가가 상위재화로 존재하며 아파트 전세가는 이에 종속적인 것으로 나타났으며, 아파트 전세는 연립주택의 매매와 단독주택의 전세로 대체될 수 있는 것으로 분석되었다. 이러한 대체효과의 공간적 패턴을 분석하기 위한 국지적 모델에서 강남구 아파트 전세가의 영향력이 주택의 종류에 따라 지역적인 차이를 보이는 것으로 나타났다. 비교적 저렴한 연립주택의 전세는 접근성이 좋지만 지가가 높은 강남구 주변 지역에 즉각적인 영향을 주는 반면, 매매는 지가가 저렴한 인천 지역에서 빠르게 반응한다. 이는 실수요를 반영하는 전세의 특성 상 대체효과가 공간적으로도 강남구의 어메니티에 대한 접근이 가능한 대체지역에서 나타남을 의미한다.

이러한 SpVAR 모델들은 각각 해석할 수 있는 바가 달라 다른 유용성을 갖고 있지만, 정확성 측면에서는 차이를 보인다. 이를 검증하고자 3장에서 제시한 MAE와 RMSE를 각 지역에서 시기별로 발생한 오차에 적용하였다. 즉, 특정 지역 i 의 t 시기의 오차를 [식 4-5]와 같이 표기했을 때, 추정의 MAE는 [식 4-6], RMSE는 [식 4-7]과 같이 구하였다. 또한 실제 미래 자료로 설정한 2015년 7월의 각 주택별 거래가격 변동량을 기초로 미래 예측의 정확성도 평가하였다. 추정된 계수를 바탕으로 2015년 6월 값을 대입했을 때 도출되는 2015년 7월의 가격 변화량을 예측값으로 하고 이를 실제 관측치와 비교하였다(식 4-8~9). 마지막으로 예측오차의 공간적 자기상관을 측정함으로써 SpVAR-Lasso모델을 통해 공간적 이질성이 필터링 되었는지를 판단하였다.

$$u_{i,t} = y_{i,t} - \hat{y}_{i,t} \quad [\text{식 4-5}]$$

$$MAE_i = \frac{1}{T} \sum_{t=1}^T |u_{i,t}| \quad [\text{식 4-6}]$$

$$RMSE_i = \sqrt{(\sum_{t=1}^T u_{i,t}^2)/T} \quad [\text{식 4-7}]$$

$$PMAE = \frac{1}{n} \sum_{i=1}^n |y_{i,2015.7} - \hat{y}_{i,2015.7}| \quad [\text{식 4-8}]$$

$$PRMSE = \sqrt{(\sum_{i=1}^n (y_{i,2015.7} - \hat{y}_{i,2015.7})^2)/n} \quad [\text{식 4-9}]$$

각 추정 방법별 오차의 비교 결과인 <표 4-11>을 보면 전체적으로 Lasso추정의 결과가 보다 정확한 것을 확인할 수 있다. 하위시장의 구획 없이 전체적인 수준에서 전역적 모델(SUR 추정)은 평균적으로 약 28만원의 오차를 보인 반면 국지적 모델(Lasso 추정)에서는 약 3만원 차이를 보여 추정의 정확도가 확연히 개선되었음을 확인할 수 있다. 이는 다른 평가 지표에서도 비슷한 패턴을 보인다. 다만, 예측 정확성은 추정 정확성에 비해 확연한 개선이라고 보기 어려운 결과를 도출하였다. 이는 추정 방식이 개선되더라도 단순히 과거와 주변 지역 가격만을 바탕으로 미래 가격을 예측하는 것의 한계를 내포하고 있다고 볼 수 있다.

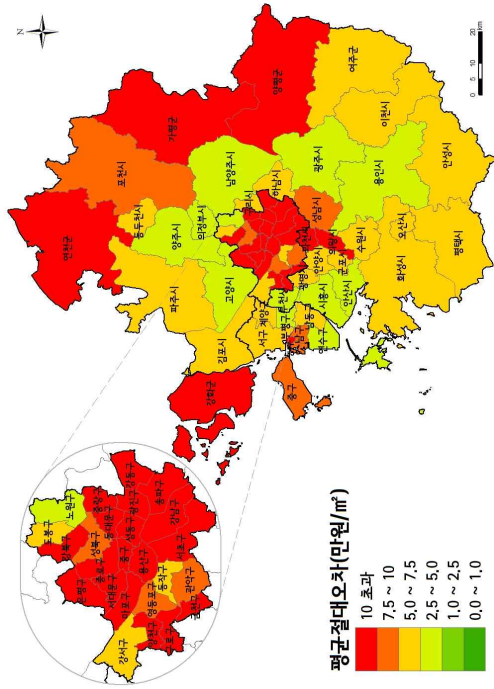
보다 구체적인 하위시장을 살펴보면, 대부분의 하위시장에서 역시 Lasso의 추정 결과가 개선된 것을 확인할 수 있다. 아파트 전세 가격의 경우, 다른 하위시장에 비해 두 가지 추정방법 모두 우수한 결과를 도출하였으나, Lasso의 추정력과 예측력이 보다 높게 나타났다. 하지만 예외적으로 단독주택의 매매시장에 대해서 SUR 추정 결과가 조금 더 정확하게 나타났지만, 이는 이상치에 크게 반응하는 평균제곱근오차의 한계로서 2015년 7월 서울특별시 중구에서 발생한 이상치에 의해 크게 왜곡된 결과라 할 수 있다(표 4-3 참조).

<표 4-11> 지역별 평균 추정 오차

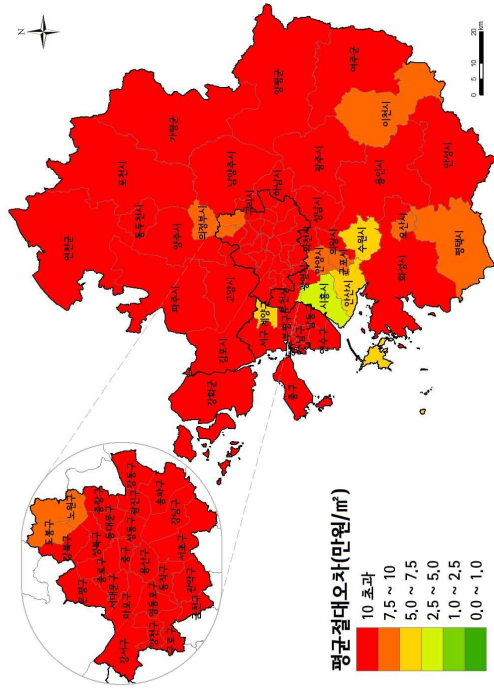
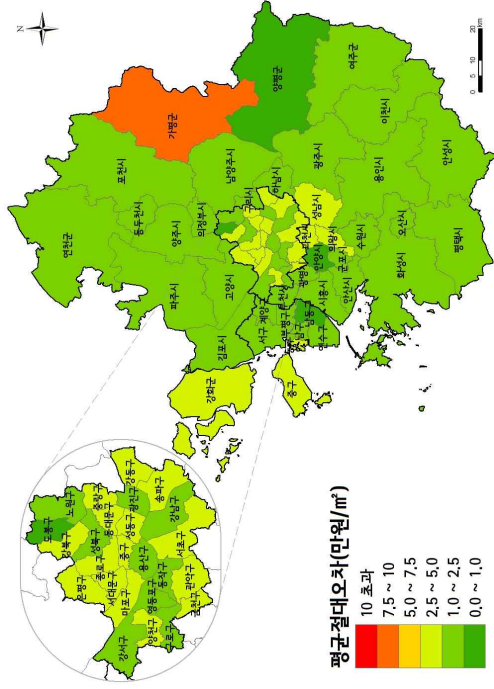
			MAE	RMSE	PMAE	PRMSE
전체		SUR	28.1746	39.2591	26.3814	49.8944
		Lasso	3.0899	18.0489	19.1183	36.8621
전세	아파트	SUR	9.4735	12.3301	18.0042	41.0034
		Lasso	2.2485	6.7756	11.9270	19.2201
	연립	SUR	15.8705	20.2613	19.7581	60.3163
		Lasso	2.3988	8.5990	13.7759	18.0929
	단독	SUR	12.4383	15.7902	13.4541	50.0813
		Lasso	2.1272	6.5961	10.3014	13.6204
매매	아파트	SUR	20.9223	28.3132	24.3074	27.0226
		Lasso	2.8909	13.1504	14.2571	22.1177
	연립	SUR	36.5517	49.2142	31.5067	31.7804
		Lasso	3.4880	20.1454	22.9230	39.4010
	단독	SUR	73.7911	109.6452	51.2576	50.2532
		Lasso	5.3861	53.0271	41.5257	72.3117

이러한 오차들의 공간적인 분포를 살펴 볼 경우 예측의 정확도가 보다 선명하게 파악된다. [그림 4-15]는 결과값의 직관적인 이해가 쉬운 평균절대오차의 공간적 분포를 지도화한 것으로 좌측이 전역적 모델, 우측이 국지적 모델의 평균절대오차 값이다. 좌측의 전역적 모델의 오차들을 살펴보면, 상대적으로 높은 오차값을 보이고 있다. 아파트 전세가의 경우 비교적 높은 정확성을 보이고 있으나, 서울 근교의 지역에서만 5만원/㎡ 이하의 오차를 보이고 있는 반면, 나머지 지역에서는 그 정확성이 떨어진다고 볼 수 있다. 아파트 매매가와 연립다세대의 전세, 매매가는 대부분의 지역에서 10만원/㎡ 이상의 오차를 보이고 있다.

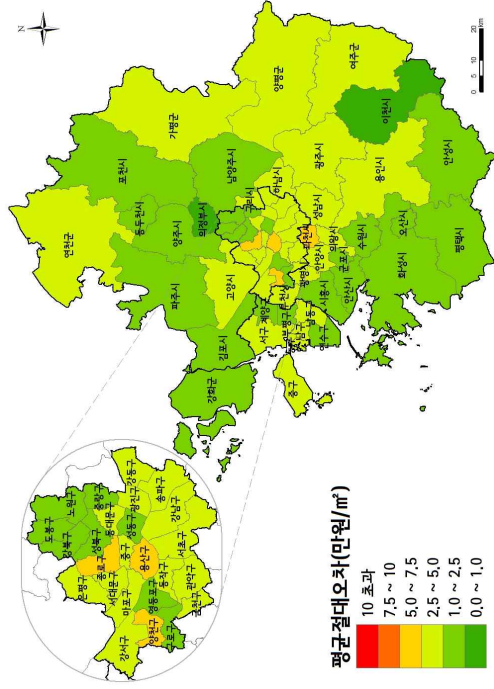
반면 우측의 Lasso 추정의 결과를 보면 대부분의 지역에서 5만원/㎡ 이하의 오차를 보이는 것을 확인할 수 있다. 특히 가격 변화가 심한 서울 지역에서의 정확성이 높아졌다. 강남지역을 예로 살펴보면 아파트 전세가의 경우 강남구에서 약 2.5만원 이하의 오차를 보이고 있으며, 아파트 매매가는 5만원 이하의 정확한 추정이 이뤄졌다. 이를 통해 국지적 모델의 결과가 주택 하위 시장의 변화를 보다 잘 포착할 수 있는 방법이라고 판단 할 수 있다.

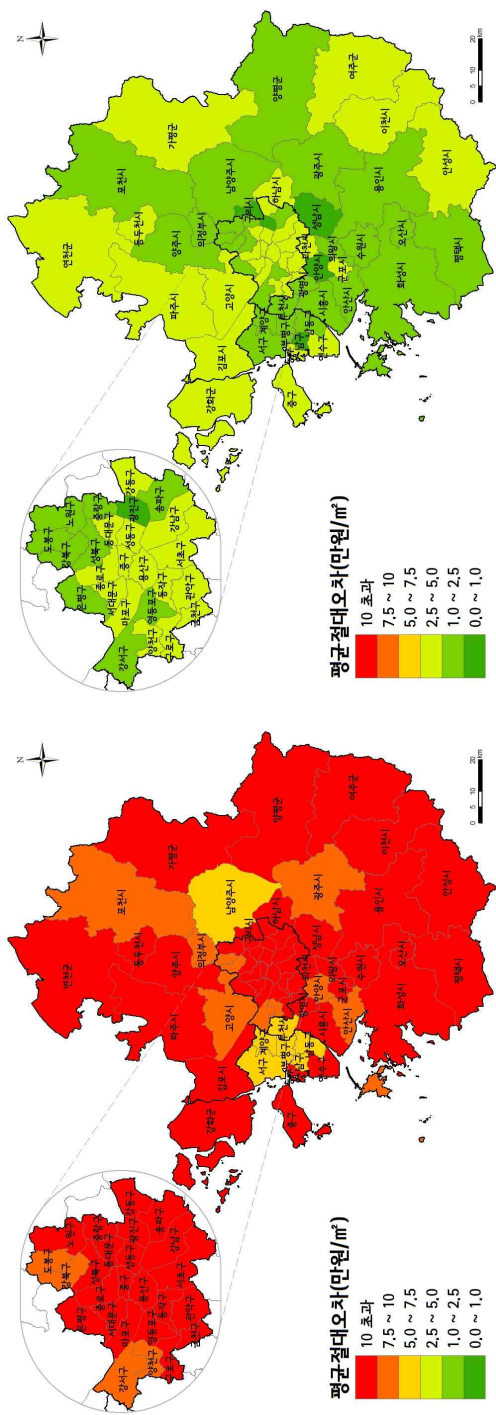


a) 아파트 전세가

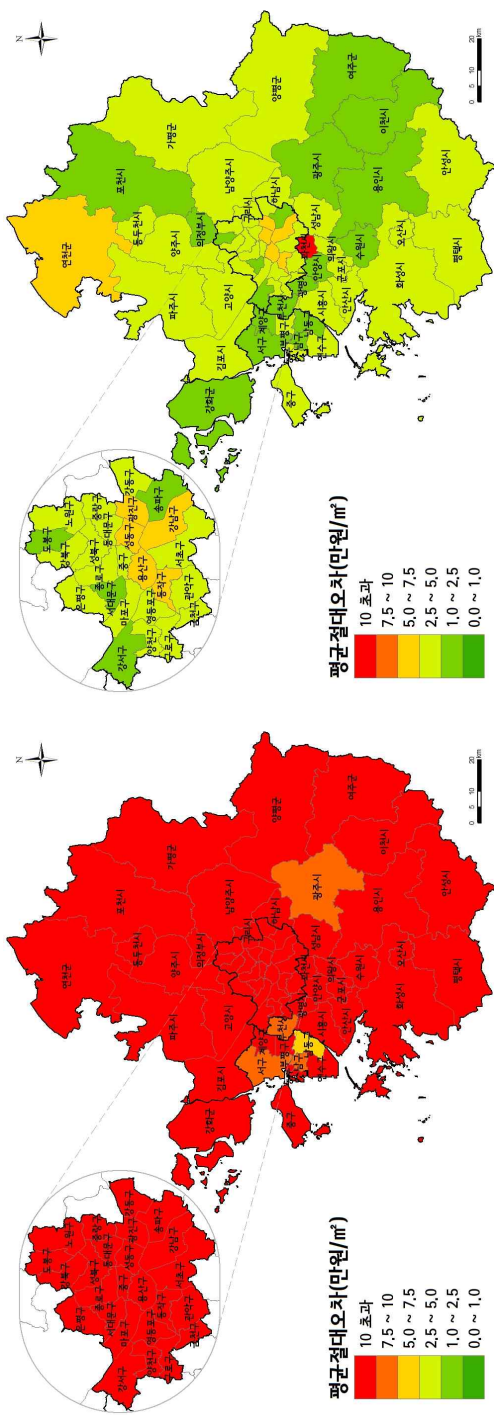


b) 아파트 매매가





c) 연립다세대 전세가



d) 연립다세대 매매가

[그림 4-15] 지역별 평균절대오차(MAE)의 공간적 분포(좌: SUR, 우: Lasso)

한편 두 추정 방법 간의 큰 차이를 보이지 않는 예측오차를 대상으로 공간적 자기상관지수를 측정하여, 오차에 발생한 공간적 자기상관을 확인하였다. 추정된 계수값을 신뢰하기 위해서는 오차항에서 상관성이 제거되어야 하기 때문이다. SUR 추정의 경우 모든 하위시장에서 통계적으로 유의한 수준에서 오차의 공간적인 자기상관이 발생하고 있으며 그 수치 또한 작다고 보기 어려우므로, 예측 결과를 충분히 신뢰하기 어렵다고 볼 수 있다(표 4-12).

<표 4-12> 예측오차의 공간적 자기상관지수(Moran's I)

변 수		Lasso	SUR
아파트	전세	-0.1607***	0.3860***
	매매	0.1938 **	0.3156***
연립 다세대	전세	-0.1350 **	0.3492***
	매매	0.0585	0.3454***
단독 다가구	전세	0.1259 **	0.2753 **
	매매	-0.0336	0.1997 **

***: 99%, **: 95%, *: 90% 신뢰구간에서 유의함

이상을 종합하면 SUR 추정을 통한 전역적 SpVAR 모델은 수도권 주택 시장의 각 주택 종류별 거래 형태에 따른 변화를 전역적인 수준에서 파악하기 좋은 방법이지만 추정과 예측의 정확성이 다소 떨어지는 방법이다. 이는 공간적으로 이질적인 주택 가격의 상호작용을 온전히 포착하지 못한 것에서 비롯되는 것으로 볼 수 있다. 또한 오차항에서 공간적인 자기상관이 발견되므로 오차항을 보정하기 위한 공간적 상관계수 행렬을 활용하는 등의 추가적인 작업이 필요하다. 반면, SpVAR-Lasso를 통한 국지적 모델은 정확성 측면에서 매우 개선된 효과를 보이고 있으며, 오차항의 상관성도 다소 해소되었다. 그러므로 SpVAR-Lasso는 공간적으로 이질성을 보이는 수도권 주택 시장에서 매매-전세가격의 시공간적 상호작용 패턴을 분석함에 있어 보다 정확하며 다양한 해석이 가능한 모델을 제시할 수 있는 방법론으로 평가할 수 있다.

제 5 장 결 론

본 연구는 주택 매매가와 전세가 간의 시공간적 상호작용 패턴을 분석하기 위한 방법을 제시하고 이를 실제 수도권 주택 시장에 적용함으로써 새로운 방법론의 활용 가능성을 탐구하였다. 기존의 주택 가격 간 시계열적인 상호작용을 분석하는 방법으로 VAR이 주로 활용되었으나, 이는 주택이 갖는 공간 효과를 다루는데 한계를 보였다. 이를 극복하고자 제시된 SpVAR은 시차항과 함께 공간적 의존성을 다루는 공간차항을 모델에 추가함으로써 주택 가격에 발생하는 시공간적 효과를 필터링할 수 있었으나, 과소표본에 의한 자유도 문제가 발생하였다. 기존에는 이러한 문제를 공간적 등질성 가정을 바탕으로 변수별로 자료를 통합함으로써 해결하고자 하여, 다수의 지역에서 발생하는 다양한 변수들의 상호작용을 파악하기 어려웠다. 이에 따라 본 연구에서는 이러한 문제점을 해결하고자 고차원 변수 중 유의한 변수 집합을 효과적으로 선택하는 방법인 Lasso를 SpVAR의 추정 방법으로 제안하였다. 이러한 SpVAR-Lasso 모델은 기존 추정 방법이 취한 자료의 통합 없이 원자료를 그대로 활용할 수 있어 정보의 손실을 피할 수 있을 뿐만 아니라 공간적으로 이질적인 상호작용 패턴을 포착하는데 유리하였다. 이러한 SpVAR-Lasso 모델을 활용하여 수도권 65개 시군구의 주택 유형별 매매-전세가의 시공간적 상호작용 패턴을 분석하고 이를 기존의 SpVAR 모델과 비교함으로써 모델의 개선 여부와 적용 가능성을 검증하였다.

보다 구체적으로 세부 연구 주제에 따른 결과를 정리하면 다음과 같다. 우선 시공간 모델인 SpVAR에 필요한 시공간적 효과의 유무를 평가하기 위한 탐색적 분석을 실시하였다. 시간적 효과를 파악하기 위해 ADF 검정과 패널 단위근 검정을 실시한 결과, 대부분의 하위시장에서 시계열적 비정상성이 입증되어 이를 제거하고자 1차 차분하여 가격변화량을 분석의 대상으로 하였다. 이러한 주택 가격의 변화량은 공간적인 효과 즉, 공간적 의존성과 공간적 이질성이 모두 있는 것으로 나타났다. Moran's I로 측정한 공간적 자기상관지수에서 아파트 전세가격은 상당히 높게 공간적으로

상호 의존적인 것으로 나타났고, 주변 지역의 다른 변수와도 밀접하게 연관이 있음이 탐지되었다. 국지적 공간자기상관지수와 지리가중회귀로 탐색한 공간적 이질성 역시 포착되어, 매매와 전세 가격의 관계가 하위시장에 따라 다르게 형성될 것임을 암시하였다. 이러한 시공간적 특성을 종합하였을 때, 주택 가격 간 상호작용 패턴을 분석하기 위해서는 다른 모델들보다 시공간 모델인 SpVAR이 적합한 것으로 나타났으며 공간적 이질성을 효과적으로 다룰 방법이 필요한 것으로 나타났다.

이에 따라 수도권 주택시장에서의 매매와 전세가격 간의 시공간적 상호작용을 보다 효과적으로 포착하기 위해 SpVAR-Lasso 모델을 제시하였다. 이는 기존의 SpVAR이 SUR 추정을 통해 계수를 추정하고 오차항에 공간 상관 행렬을 추가하는 복잡한 과정을 벗어난 것으로, 다수의 변수를 그대로 활용하여 유의한 변수 집합을 선택할 수 있었다. 이를 통해 도출한 국지적 모델을 활용하여 강남구의 아파트 가격의 대체효과를 살펴본 결과, 전역적 모델과 달리 국지적 모델에서는 강남구의 아파트 매매가와 전세가 간에는 대체효과가 발생하지 않았다. 또한 매매가격이 아닌 전세가격이 다른 하위시장 가격에 영향을 받지 않고 독립적으로 존재하는 것으로 나타났다. 대신 강남구 아파트 전세가는 상급재화로서 서울 외곽지역이나 근교의 연립다세대 혹은 단독다가구 주택의 전세가에 영향을 주었다. 이를 보다 명확히 살피기 위해 실시한 충격반응분석의 결과, 외부적 호재에 의해 강남구 아파트 전세가격이 상승하게 될 경우 아파트 가격은 시차를 두고 경기 외곽까지 영향을 미치지만 이러한 영향은 조정 작용을 통해 4기 이내에서 대부분 소멸되었다. 또한 연립주택의 경우 비교적 가격이 저렴한 전세는 접근성이 좋은 주변지역으로 대체효과가 빠르게 나타났지만, 매매는 인천 지역 등 비교적 원거리에서 나타났다. 하지만 이러한 변화는 분산 분석 결과 해당 하위시장의 가격 변화에 영향이 크지 않은 것으로 나타났다. 이는 다양한 변수에 의해 결정되는 주택 가격을 다른 하위시장의 가격만으로 설명하고자 하는데서 발생하는 한계라고 볼 수 있다.

마지막으로 이렇게 도출된 SpVAR-Lasso를 기존의 SpVAR 모델과 비교함으로써 주택 가격 연구에서의 적용 가능성을 모색해보았다. 추정과 예

측의 정확성을 MAE와 RMSE를 통해 평가한 결과, 대부분의 지표에서 국지적 모델인 SpVAR-Lasso에서 보다 정확한 추정값과 예측값을 도출할 수 있었으며 실제 관측값과 매우 유사한 정도의 결과를 제시하였다. 또한 오차항의 공간적 자기상관을 측정함으로써 모델의 시공간적 효과의 필터링 여부를 살펴보았다. 기존의 SpVAR 추정 방법은 공간적 효과의 하나인 공간적 이질성을 다루지 못하여 오차항에서 공간적 자기상관이 유의한 수준으로 포착된 반면, SpVAR-Lasso 모델에서는 이러한 문제가 해결되었다. 이상을 종합해보면, 비록 모델의 해석상에 현상의 인과관계에 대한 설명이 부족하다는 한계를 내포하지만 기존 모델에 비해 정확하며 공간적으로 다양한 해석이 가능하다는 점에서 SpVAR-Lasso 모델은 주택 가격의 상호작용 패턴 분석에 있어 충분히 활용될 수 있는 방법으로 볼 수 있다.

그러므로 본 연구의 의의는 다음과 같다. 첫째, 기존의 연구에서 비교적 단순하게 구획한 주택 시장을 두 가지 기준으로 나누고 이를 또 매매와 전세시장으로 나누어 살펴보으로써 보다 구체적이고 하위시장 정의에 가까운 분석 단위를 설정하였다는 점이다. 이를 통해 기존 연구 결과들에 비해 심도 있고 다양한 해석이 가능한 분석이 이루어졌다. 이는 기존의 방법에 비해 고차원 자료를 다룸에 있어 효과적인 방법론의 개선을 통해 시도할 수 있었다. 구체적으로 시공간 모델링 기법인 SpVAR이 갖는 과소표본에 의한 자유도의 문제를 축소추정법의 일종인 Lasso를 이용하여 해결했다는 점이다. 이는 관측값보다 변수가 많은 경우에도 모델을 활용할 수 있게 하여, 세분화 된 주택 하위시장을 분석하기에 적합하였다. 또한 이를 통해 주택 시장이 갖는 공간적 이질성을 반영하는 국지적 모델링이 가능하였고, 기존 방법에 비해 보다 정확한 결과를 도출할 수 있었다. 마지막으로 주택 시장의 세분화와 다변량 추정에 용이한 국지적 SpVAR 모델을 통해 수도권 주택 시장에서의 매매와 가격 간 상호작용을 전역적 수준과 국지적 수준으로 구분하여 다각도로 분석하였다는 점이다. 특히, 국지적 수준에서의 분석은 기존의 연구들에서는 잘 다루어지지 않았던 주택 가격 간 시공간적 대체효과를 밝혔다는 점에서 본 연구는 기존의 연구와 차별성을 갖는다.

추가적으로 본 연구는 다음과 같은 연구들을 통해 보완되고 발전될 수 있을 것이다. 우선 SpVAR-Lasso 모델의 성능에 대한 일반적인 수준에서의 확인이 필요하다. 개별 사례인 수도권 주택 가격에 대한 분석에서는 기존의 방법에 비해 정확한 결과를 도출하였으나, 이것이 우연의 결과인지 아닌지를 다양한 시뮬레이션을 통해 검증해볼 필요가 있다. 또한 모델의 해석의 측면에서 보다 다양한 지역 간 매매-전세가격의 상호작용을 살펴볼 필요가 있다. 본 연구에서는 주택 가격이 높은 강남구의 아파트를 중점적으로 살펴보았지만, 변화율이 심한 하위시장에서의 변화가 다른 주택시장에 대해 어떠한 영향력을 갖는지 분석함으로써 수도권 주택시장 가격 변화에 대한 보다 폭넓은 이해가 가능할 것이다. 마지막으로 주택가격 간 시공간적 상호작용의 원인에 대해 탐구하는 것도 필요하다. 본 연구에서는 강남구와 강서구의 아파트 전세가격 간 대체관계를 밝히고는 있으나 그 원인에 대해서 추측할 뿐 명백한 근거는 본 연구를 통해 제시할 수 없었다. 그러므로 이를 추가적으로 분석하는 연구가 수행된다면, 본 연구가 밝혀낸 주택시장에서의 매매와 전세가격의 상호작용 현상이 보다 풍부하게 해석될 것이다.

참 고 문 헌

- 강영옥, 2004, “주택가격 결정인자의 공간적 다양성 모델링,” 대한지리학회지, 39(6), 907-921.
- 강창덕, 2010, “GWR 접근법을 활용한 부동산 감정평가 모형 연구 : 서울시 아파트를 사례로,” 부동산연구, 20(2), 107-132.
- 국토교통부, 2014, 2014년도 주거실태조사.
- 김경민, 2007, “강남지역의 아파트가격 변화가 전국에 미치는 영향,” 국토계획, 42(2), 137-152.
- 김경환 · 서승환 · 유진방, 1991, “우리나라 부동산 가격과 물가에 관한 실증분석,” 금융경제연구소.
- 김대원 · 유정석, 2015, “서울시 아파트 전세가격 변동률 및 결정요인의 공간 전이 효과 분석,” 부동산학연구, 21(1), 31-55.
- 김성우 · 정건섭, 2010, “부산 아파트 실거래가를 이용한 전통적 헤도닉모형과 공간계량모형간의 적합도에 관한 비교 연구,” 부동산학연구, 16(3), 41-55.
- 김소연 · 김영호, 2013, “주거지 인문환경의 공간 속성을 고려한 주택 가격 결정 모델,” 한국지도학회지, 13(3), 41-56.
- 김영호, 2011, “공간네트워크의 이변량공간상관관계를 이용한 서울시 자전거와 버스 대중교통의 연계 가능성 분석,” 한국도시지리학회지, 14(3), 55-72.
- 김의준 · 김양주 · 신명수, 2000, “수도권 아파트 가격의 지역간 인과성 분석,” 국토계획, 35(4), 109-117.
- 김태호 · 이창무, 2006, “그린벨트 및 주택의 어메니티요소가 주택임대료에 미치는 영향력의 시계열적 변화,” 국토계획, 41(5), 61-79.
- 김현재, 2003, “서울시 아파트 매매 및 전세가격 결정요인의 분석,” 부동산학보, 22, 98-121.
- 김호철 · 최창규, 2008, “수도권 재건축아파트와 일반아파트 가격간의 인과성 및 지역 파급 효과에 대한 실증적 분석” 한국지역개발학회

- 지, 20(1), 39-58.
- 문권순, 1997, “벡터자기회귀(VAR)모형의 이해,” 통계분석연구, 2(1), 23-56.
- 박동국 · 천인호, 2006, “구조적 벡터자기회귀(SVAR)를 이용한 서울지역 아파트 매매, 전세가격의 상관관계,” 한일경상논집, 35, 149-174.
- 박창이 · 김용대 · 김진석 · 송종우 · 최호식, 2011, R을 이용한 데이터마 이닝, 교우사.
- 박헌수 · 김태경, 2008, “부동산가격에 있어 장기균형과 충격반응분석,” 국토계획, 43(5), 35-48.
- 박헌수, 2003, “시공간자기회귀모형을 이용한 서울 아파트 가격의 추정,” 국토연구, 38, 95-106.
- 박헌수 · 유은영, 2014, “공간패널모형을 활용한 우리나라 주택가격의 동 학적 특성분석,” 지역연구, 30(1), 3-18.
- 서승환, 2007, “주택가격 변화의 지역연관성에 관한 연구,” 서울도시연구, 8(4), 1-13.
- 손재영, 2000, “주택매매가격과 전세가격의 상관관계,” 사회과학논총, 24, 139-163.
- 손철, 2011, “수도권의 공간적 주택하위시장 식별에 대한 연구,” 국토연 구, 70, 151-166.
- 신종협 · 서대교, 2014, “국내 주택시장의 동태적 상관관계 분석,” 부동산 연구, 24(1), 15-26.
- 신현중, 2006, “정책분석 및 평가에서 통합시계열자료를 사용할 경우에 패널단위근 검정의 필요성에 대한 고찰,” 사회과학연구, 22(2), 341-362.
- 원제무 · 정광섭 · 김상원 · 백진호 · 백기영, 2009, “시장세분화를 통한 주상복합주택 가격 결정요인 특성에 관한 연구,” 국토계획, 44(3), 137-147.
- 윤종인, 2015, “주택 매매가격과 전세가격의 비대칭적 움직임에 대한 실 증연구,” 경영컨설팅연구, 15(2), 211-220.

- 이변송 · 정의철 · 김용현, 2002, “아파트 단지특성이 아파트 가격에 미치는 영향에 관한 연구,” 국제경영연구, 8(2), 21-45.
- 이영수, 2008, “한국의 주택가격과 거시경제,” 부동산학연구, 14(3), 129-147.
- 이용만 · 이상한, 2004, “강남지역의 주택가격이 주변지역의 주택가격을 결정하는가?,” 국토계획, 39(1), 73-91.
- 이용만, 2000, “구조적 변화인가 가격상승의 징조인가,” 부동산학연구, 6(1), 9-22.
- 이인화 · 문영기, 2007, “공동주택단지의 심미적 디자인요인이 아파트가격 형성에 미치는 영향,” 주택연구, 15(3), 169-194.
- 임규채 · 기석도, 2006, “주택시장의 전세가격과 매매가격간의 상호관계에 관한 연구,” 산업경제연구, 19(3), 1203-1223.
- 임정호, 2006, “주택매매시장, 전세시장 및 월세시장 간의 상호연관성에 관한 연구,” 주택연구, 14(1), 165-193.
- 임재만, 2004, “서울지역 아파트 매매시장과 전세시장의 관계에 관한 연구,” 감정평가연구, 14(2), 163-177.
- 임재만, 2013, “주거이력구조를 고려한 매매가격과 전세가격의 관계,” 서울도시연구, 14(2), 35-50.
- 전경구, 2012, “교육환경이 아파트 가격에 미치는 상호작용 효과분석-위계선형모형을 중심으로,” 한국지역개발학회지, 24(3), 1-24.
- 전해정, 2012, “주택 매매시장, 전세시장, 월세시장간의 동학적 관계에 대한 연구,” 지역연구, 28(3), 123-139.
- 전해정 · 박현수, 2014, “시공간자기회귀(STAR)모형을 이용한 부동산 가격 추정에 관한 연구,” 부동산연구, 24(1), 7-14.
- 정규일, 2006, “자산가격과 유동성간의 관계분석,” 한국경제연구, 17, 257-287.
- 조주현 · 임정호, 2004, “전세가격과 매매가격 및 월세가격간의 관계에 관한 연구,” 부동산학연구, 10(2), 17-29.
- 진찬우 · 이건학, 2014, “GWL을 적용한 공간 헤도닉 모델링,” 대한지리학

- 회지, 49(6), 917-934.
- 최두열, 2012, “서울시 구별 전세가의 수렴과 그 결정 요인에 관한 연구,”
서울도시연구, 13(1), 57-76.
- 최현일 · 임병진, 2011, “APT 전세가와 매매가 상호 영향력에 관한 실증
적 연구,” 대한경영학회지, 24(6), 3707-3722.
- 황두현, 1990, “주택매매가격과 전세가격의 시차분석,” 경제연구, 6, 23-38.
- Abraham J. M., and P.H. Hendershott, 1996, Bubbles in
Metropolitan Housing Markets, *Journal of Housing
Research*, 7(2), 191-207.
- Allen, M. T., Springer, T. M., and Waller, N. G., 1995, Implicit
pricing across residential rental submarkets, *The Journal
of Real Estate Finance and Economics*, 11(2), 137-151.
- Anari, A., and Kolari, J., 2002, House prices and inflation, *Real
Estate Economics*, 30(1), 67-84.
- Anselin, L., 1988, *Spatial econometrics: methods and models*,
Springer Science & Business Media.
- Anselin, L., 1995, Local indicators of spatial association-LISA,
Geographical analysis, 27(2), 93-115.
- Anselin, L., and Lozano-Gracia, N., 2008, Errors in variables and
spatial effects in hedonic house price models of ambient
air quality, *Empirical economics*, 34(1), 5-34.
- Aoki, K., Proudman, J., and Vlieghe, G., 2002, Houses as collateral:
has the link between house prices and consumption in
the UK changed?, *Federal Reserve Bank of New York
Economic Policy Review*, 8(1), 163-78.
- Bailey, T. C., and Gatrell, A. C., 1995, *Interactive spatial data
analysis*, Essex: Longman Scientific & Technical.
- Bajic, V., 1985, Housing-market segmentation and demand for
housing attributes: Some empirical findings, *Real Estate*

- Economics*, 13(1), 58-75.
- Ball, M. J., and Kirwan, R. M., 1977, Accessibility and supply constraints in the urban housing market, *Urban studies*, 14(1), 11-32.
- Bárcena, M. J., Menéndez, P., Palacios, M. B., and Tusell, F., 2014, Alleviating the effect of collinearity in geographically weighted regression, *Journal of Geographical Systems*, 16(4), 441-466.
- Beenstock, M., and Felsenstein, D., 2007, Spatial vector autoregressions, *Spatial Economic Analysis*, 2(2), 167-196.
- Beenstock, M., and Felsenstein, D., 2010, Spatial error correction and cointegration in nonstationary panel data: regional house prices in Israel, *Journal of Geographical Systems*, 12(2), 189-206.
- Bourassa, S. C., Hamelink, F., Hoesli, M., and MacGregor, B. D., 1999, Defining housing submarkets, *Journal of Housing Economics*, 8(2), 160-183.
- Bourassa, S. C., Hoesli, M., and Peng, V. S., 2003, Do housing markets really matter?, *Journal of Housing Economics*, 12(1), 12-28.
- Box, G. E., and Jenkins, G. M., 1976, *Time series analysis: forecasting and control*, John Wiley & Sons.
- Brunsdon, C., Fotheringham, S., and Charlton, M., 1998, Geographically weighted regression, *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3), 431-443.
- Campbell, S. D., Davis, M. A., Gallin, J., and Martin, R. F., 2009, What moves housing markets: A variance decomposition of the rent-price ratio, *Journal of Urban Economics*, 66(2),

90-102.

- Can, A., 1992, Specification and estimation of hedonic housing price models, *Regional science and urban economics*, 22(3), 453-474.
- Chen, M. C., and Patel, K., 1998, House price dynamics and Granger causality: an analysis of Taipei new dwelling market, *Journal of the Asian Real Estate Society*, 1(1), 101-126.
- Cressie, N., 1993, *Statistics for spatial data*. John Wiley & Sons.
- Cressie, N., and Wikle, C. K., 2011, *Statistics for spatio-temporal data*, John Wiley & Sons.
- Davis, M. A., and Heathcote, J., 2005, Housing and the business cycle, *International Economic Review*, 46(3), 751-784.
- Dickey, D. A., and Fuller, W. A., 1979, Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American statistical association*, 74(366a), 427-431.
- Dubé, J., and Legros, D., 2013, A spatio-temporal measure of spatial dependence: An example using real estate data, *Papers in Regional Science*, 92(1), 19-30.
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R., 2004, Least angle regression, *The Annals of statistics*, 32(2), 407-499.
- Feldstein, M., 1980, Inflation, tax rules, and the prices of land and gold, *Journal of Public Economics*, 14(3), 309-317.
- Fotheringham S., Brunsdon C., Charlton M., 2002, *Geographically weighted regression*, Wiley.
- Fotheringham, A. S., Crespo, R., and Yao, J., 2015a, Exploring, modelling and predicting spatiotemporal variations in house prices, *The Annals of Regional Science*, 54(2),

417-436.

- Fotheringham, A. S., Crespo, R., and Yao, J., 2015b, Geographical and Temporal Weighted Regression (GTWR), *Geographical Analysis*, 1-22.
- Friedman, J., Hastie, T., and Tibshirani, R., 2010, Regularization paths for generalized linear models via coordinate descent, *Journal of statistical software*, 33(1), 1.
- Gallin, J., 2008, The Long-Run Relationship Between House Prices and Rents, *Real Estate Economics*, 36(4), 635-658.
- Geary, R. C., 1954, The contiguity ratio and statistical mapping, *The incorporated statistician*, 5(3), 115-146.
- Giuliodori, M., 2005, The role of house prices in the monetary transmission mechanism across European countries, *Scottish journal of political economy*, 52(4), 519-543.
- Giussani, B., and Hadjimatheou, G., 1991, Modeling regional house prices in the United Kingdom, *Papers in Regional Science*, 70(2), 201-219.
- Goodman, A. C., 1978, Hedonic prices, price indices and housing markets. *Journal of Urban Economics*, 5(4), 471-484.
- Goodman, A. C., and Thibodeau, T. G., 2003, Housing market segmentation and hedonic prediction accuracy, *Journal of Housing Economics*, 12(3), 181-201.
- Granger C. W., 1969, Investigating causality relations by econometric models and cross-special methods, *Econometrica*, 37(3), 424-438.
- Gujarati, D. N., and Porter, D. C., 2012, *Basic econometrics*, Tata McGraw-Hill Education.
- Harrison, D., and Rubinfeld, D. L., 1978, Hedonic housing prices and the demand for clean air, *Journal of environmental*

- economics and management*, 5(1), 81-102.
- Hoerl, A. E., and Kennard, R. W., 1970, Ridge regression, Biased estimation for nonorthogonal problems, *Technometrics*, 12(1), 55-67.
- Hsu, N. J., Hung, H. L., and Chang, Y. M., 2008, Subset selection for vector autoregressive processes using lasso. *Computational Statistics & Data Analysis*, 52(7), 3645-3657.
- Huang, B., Wu, B., and Barry, M., 2010, Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24(3), 383-401.
- Hyndman, R. J., and Koehler, A. B., 2006, Another look at measures of forecast accuracy, *International journal of forecasting*, 22(4), 679-688.
- Im, K. S., Pesaran, M. H., and Shin, Y., 2003, Testing for unit roots in heterogeneous panels, *Journal of econometrics*, 115(1), 53-74.
- Julliard, C., and Wong, G., 2008, International Comparison of Housing Risk Premia, *Wharton Business School V Working Paper*.
- Kan, K., Kwong, S. K. S., and Leung, C. K. Y., 2004, The Dynamics and Volatility of Commercial and Residential Property Prices: Theory and Evidence. *Journal of Regional Science*, 44(1), 95-123.
- Kinal, T., and Ratner, J., 1986, A VAR forecasting model of a regional economy: its construction and comparative accuracy. *International Regional Science Review*, 10(2), 113-126.
- Kuethé, T. H., and Pede, V. O., 2011, Regional housing price cy-

- cles: a spatio-temporal analysis using US state-level data, *Regional studies*, 45(5), 563-574.
- Lastrapes, W. D., 2002, The real price of housing and money supply shocks: time series evidence and theoretical simulations, *Journal of Housing Economics*, 11(1), 40-74.
- Lee, G., Cho, D., and Kim, K., 2015, The modifiable areal unit problem in hedonic house-price models, *Urban Geography* (in press).
- Lee, S. I., 2001, Developing a bivariate spatial association measure: an integration of Pearson's r and Moran's I , *Journal of Geographical Systems*, 3(4), 369-385.
- Leung, C., 2004, Macroeconomics and housing: a review of the literature, *Journal of Housing Economics*, 13(4), 249-267.
- Levin, A., Lin, C. F., and Chu, C. S. J., 2002, Unit root tests in panel data: asymptotic and finite-sample properties, *Journal of econometrics*, 108(1), 1-24.
- Lowry, I. S., Hillestad, C. E., and Sarma, S., 1983, *California's Housing*, Rand Corporation.
- Meen, G., 1999, Regional house prices and the ripple effect: a new interpretation, *Housing studies*, 14(6), 733-753.
- Moran, P. A., 1950, Notes on continuous stochastic phenomena, *Biometrika*, 37(1), 17-23.
- Munro, M., 1986, Testing for segmentation in the private housing market in Glasgow, *Centre for Housing Research*, Discussion Paper No 8.
- Pace, R. K., Barry, R., Clapp, J. M., and Rodriguez, M., 1998, Spatiotemporal autoregressive models of neighborhood effects, *The Journal of Real Estate Finance and Economics*, 17(1), 15-33.

- Parent, O., and Vom Hofe, R., 2013, Understanding the impact of trails on residential property values in the presence of spatial dependence, *The Annals of Regional Science*, 51(2), 355-375.
- Phillips, R. S., 1988, Unravelling the residential rent-value puzzle, *Urban Studies*, 25(6), 487-496.
- Pollakowski, H. O., and Ray, T. S., 1997, Housing price diffusion patterns at different aggregation levels: an examination of housing market efficiency, *Journal of Housing Research*, 8(1), 107-124.
- Poterba, J. M., Weil, D. N., and Shiller, R., 1991, House price dynamics: The role of tax policy and demography, *Brookings Papers on Economic Activity*, 1991(2), 143-203.
- Quah, D., 1994, Exploiting cross-section variation for unit root inference in dynamic data, *Economics letters*, 44(1), 9-19.
- Rosen, S., 1974, Hedonic prices and implicit markets: Product differentiation in pure competition, *Journal of Political Economy*, 82(1), 34-55.
- Rosen, H. S., 1979, Housing decision and US income tax, *Journal of Public Economics*, 11(1), 1-23.
- Schnare, A. B., and Struyk, R. J., 1976, Segmentation in urban housing markets, *Journal of Urban Economics*, 3(2), 146-166.
- Sims, C. A., 1980, Macroeconomics and reality, *Econometrica*, 48(1) 1-48.
- Sonstelie, J. C., and Portney, P. R., 1980, Gross rents and market values: testing the implications of Tiebout's hypothesis, *Journal of Urban Economics*, 7(1), 102-118.
- Stein, J. C., 1995, Prices and Trading Volume in the Housing

- Market: A Model with Down-Payment Effects, *The Quarterly Journal of Economics*, 110(2), 379-406.
- Sun, H., Tu, Y., and Yu, S. M., 2005, A Spatio-Temporal Autoregressive Model for Multi-Unit Residential Market Analysis, *The Journal of Real Estate Finance and Economics*, 31(2), 155-187.
- Swan, C., 1984, A model of rental and owner-occupied housing, *Journal of Urban Economics*, 16(3), 297-316.
- Tibshirani, R., 1996,. Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267-288.
- Titman, S., 1982, The effects of anticipated inflation on housing market equilibrium, *The Journal of Finance*, 37(3), 827-842.
- Tsatsaronis, K., and Zhu, H., 2004, What drives housing price dynamics: cross-country evidence, *BIS Quarterly Review*, March, 65-78.
- Watkins, C. A., 2001, The definition and identification of housing submarkets, *Environment and Planning A*, 33(12), 2235-2253.
- Wheeler, D. C., 2009, Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted lasso. *Environment and planning. A*, 41(3), 722.
- Zellner, A., 1962, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of the American statistical Association*, 57(298), 348-368.
- Zhu, J., and Hastie, T., 2004, Classification of gene microarrays by penalized logistic regression, *Biostatistics*, 5(3), 427-443.

| ABSTRACT |

Analysis of the Spatio-temporal Interaction Patterns of the Housing Sales and Rent Price Using SpVAR-Lasso

Jin, Chanwoo

Department of Geography

Graduate School

Seoul National University

This thesis aims to propose a new model for analysis of spatio-temporal interaction patterns of sales and rent price in housing market, and apply the model to the Seoul Metropolitan Area for evaluation of its applicability. Locally heterogeneous interaction patterns are focused because the existing spatio-temporal model, SpVAR, is inappropriate due to insufficient degree of freedom. Therefore, Lasso, which is free from the ‘small n problem’, is suggested as a new estimation method for SpVAR model(SpVAR-Lasso) in order to overcome the limitations and capture heterogeneous interaction patterns of diverse housing submarkets. Also a process of SpVAR-Lasso, data exploration - modeling - application and evaluation, is proposed and applied to the case of 65 Sigungu in Seoul Metropolitan Area for the last 5 years to evaluate its applicability and performance. The results of this case study are summarized as follows.

In the exploratory approach, spatial and temporal effects of house price data are measured and evaluated for the application of SpVAR. The temporal effects of the data are proven to be non-stationary, so the housing prices need first-differentiating. In this variation of housing prices, the spatial effects such as spatial dependence and spatial heterogeneity, are also found.

Apartment rental prices show particularly high relative spatial autocorrelation and spatial correlation with other variables in neighborhoods. In addition, spatial heterogeneity, which indicates different relations to variables depending on the region, is also detected. Therefore, SpVAR-Lasso model is a proper method to analyze the spatio-temporal interactions of prices in housing submarkets because it is an effective filter for spatio-temporal dependence effects and controls the spatial heterogeneity.

Thus, spatio-temporal interactions of house sales and rent prices in Seoul Metropolitan Area are locally modeled by SpVAR-Lasso. When examining the substitution effect of apartment prices of Gangnam-gu, rent prices present independence from other goods unlike the global model. Instead, rental prices of apartment in Gangnam-gu affect to multi-family or single housing in suburbs of Seoul as a superior goods. If the prices rise due to an external factor, prices of apartment in other regions are influenced, but by locally different temporal lags whose impacts fade away after 4 months because of feedback effects. In addition, in the case of multi-family housing rent which is relatively inexpensive, the substitution effect occurs to the neighborhood where has close proximity, while sales occurs in the region where is relatively far such as Incheon.

Finally, SpVAR-Lasso is compared with SpVAR to evaluate applicability of the new model in the study on housing price. As a result, estimated and predicted values of SpVAR-Lasso are much closer to observations than SpVAR. Furthermore, in terms of filtering of spatial and temporal effects, the new model shows lower spatial autocorrelation of errors. Hence, it is possible to say that SpVAR-Lasso is applicable to analyze spatio-temporal interactions of housing prices because it has improvement on accuracy and can interpret the regions in detail.

The contribution of this thesis is to capture local spatio-temporal interaction patterns of diverse housing prices by suggesting Lasso as an estimation method for SpVAR model. Additionally, spatio-temporal substitution effects in housing prices are identified empirically by applying the new improved model to a real case, Seoul Metropolitan Area.

Keywords: SpVAR, Lasso, SpVAR-Lasso, Spatial Dependence, Spatial Heterogeneity, Spatio-temporal Interaction, Housing Sales-Rent Price, Housing Market in Seoul Metropolitan Area

Student Number: 2014-20248